
Pending Nucleic Acid and/or Pending Amino Acid database searches now generate two sets of results. These databases were split into two parts to reduce the time needed to update the databases daily. The split freed up more machine time for processing searches.

Searches run against the Nucleic Acid Pending database produce two sets of results, with the extensions, **.rnpm** and **.rnpn**

Searches run against the Amino Acid Pending database produce two sets of results, with the extensions, **.rapm** and **.rapn**

The Pending database search results should not be left in the case because they contain data that is confidential.

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 19, 2002, 07:21:40 ; Search time 5830.82 seconds
(without alignments)
6486.799 Million cell updates/sec

Title: US-09-763-292-1

Perfect score: 1746

Sequence: 1 cttcaatggatccctttgtg.....acatattattataaataaga 1746

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 22023303 seqs, 10831430700 residues

Total number of hits satisfying chosen parameters: 44045606

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending_Patents_NA_Main:*

```
1: /cgn2_6/ptodata/2/pna/PCTUS_COMB.seq:*
2: /cgn2_6/ptodata/2/pna/US06_COMB.seq:*
3: /cgn2_6/ptodata/2/pna/US07_COMB.seq:*
4: /cgn2_6/ptodata/2/pna/US080_COMB.seq:*
5: /cgn2_6/ptodata/2/pna/US081_COMB.seq:*
6: /cgn2_6/ptodata/2/pna/US082_COMB.seq:*
7: /cgn2_6/ptodata/2/pna/US083_COMB.seq:*
8: /cgn2_6/ptodata/2/pna/US084_COMB.seq:*
9: /cgn2_6/ptodata/2/pna/US085_COMB.seq:*
10: /cgn2_6/ptodata/2/pna/US086_COMB.seq:*
11: /cgn2_6/ptodata/2/pna/US087_COMB.seq:*
12: /cgn2_6/ptodata/2/pna/US088_COMB.seq:*
13: /cgn2_6/ptodata/2/pna/US089_COMB.seq:*
14: /cgn2_6/ptodata/2/pna/US090_COMB.seq:*
15: /cgn2_6/ptodata/2/pna/US091_COMB.seq:*
16: /cgn2_6/ptodata/2/pna/US092_COMB.seq:*
17: /cgn2_6/ptodata/2/pna/US093_COMB.seq:*
18: /cgn2_6/ptodata/2/pna/US094_COMB.seq:*
19: /cgn2_6/ptodata/2/pna/US095A_COMB.seq:*
20: /cgn2_6/ptodata/2/pna/US095B_COMB.seq:*
21: /cgn2_6/ptodata/2/pna/US095C_COMB.seq:*
22: /cgn2_6/ptodata/2/pna/US096A_COMB.seq:*
23: /cgn2_6/ptodata/2/pna/US096B_COMB.seq:*
24: /cgn2_6/ptodata/2/pna/US096C_COMB.seq:*
25: /cgn2_6/ptodata/2/pna/US096D_COMB.seq:*
26: /cgn2_6/ptodata/2/pna/US096E_COMB.seq:*
27: /cgn2_6/ptodata/2/pna/US096F_COMB.seq:*
28: /cgn2_6/ptodata/2/pna/US097A_COMB.seq:*
29: /cgn2_6/ptodata/2/pna/US097B_COMB.seq:*
30: /cgn2_6/ptodata/2/pna/US097C_COMB.seq:*
31: /cgn2_6/ptodata/2/pna/US098_COMB.seq:*
32: /cgn2_6/ptodata/2/pna/US099_COMB.seq:*
33: /cgn2_6/ptodata/2/pna/US6000_COMB.seq:*
34: /cgn2_6/ptodata/2/pna/US6001_COMB.seq:*
35: /cgn2_6/ptodata/2/pna/US6002_COMB.seq:*
36: /cgn2_6/ptodata/2/pna/US6003_COMB.seq:*
37: /cgn2_6/ptodata/2/pna/US6004_COMB.seq:*
38: /cgn2_6/ptodata/2/pna/US6005_COMB.seq:*
39: /cgn2_6/ptodata/2/pna/US6006_COMB.seq:*
40: /cgn2_6/ptodata/2/pna/US6007_COMB.seq:*
41: /cgn2_6/ptodata/2/pna/US6008_COMB.seq:*
42: /cgn2_6/ptodata/2/pna/US6009_COMB.seq:*
43: /cgn2_6/ptodata/2/pna/US6010_COMB.seq:*
```

```
44: /cgn2_6/ptodata/2/pna/US6011_COMB.seq:*
45: /cgn2_6/ptodata/2/pna/US6012_COMB.seq:*
46: /cgn2_6/ptodata/2/pna/US6013_COMB.seq:*
47: /cgn2_6/ptodata/2/pna/US6014_COMB.seq:*
48: /cgn2_6/ptodata/2/pna/US6015_COMB.seq:*
49: /cgn2_6/ptodata/2/pna/US6016_COMB.seq:*
50: /cgn2_6/ptodata/2/pna/US6017_COMB.seq:*
51: /cgn2_6/ptodata/2/pna/US6018_COMB.seq:*
52: /cgn2_6/ptodata/2/pna/US6019_COMB.seq:*
53: /cgn2_6/ptodata/2/pna/US6020_COMB.seq:*
54: /cgn2_6/ptodata/2/pna/US6021_COMB.seq:*
55: /cgn2_6/ptodata/2/pna/US6022_COMB.seq:*
56: /cgn2_6/ptodata/2/pna/US6023_COMB.seq:*
57: /cgn2_6/ptodata/2/pna/US6024_COMB.seq:*
58: /cgn2_6/ptodata/2/pna/US6025_COMB.seq:*
59: /cgn2_6/ptodata/2/pna/US6026_COMB.seq:*
60: /cgn2_6/ptodata/2/pna/US6027_COMB.seq:*
61: /cgn2_6/ptodata/2/pna/US6028_COMB.seq:*
62: /cgn2_6/ptodata/2/pna/US6029_COMB.seq:*
63: /cgn2_6/ptodata/2/pna/US6030_COMB.seq:*
64: /cgn2_6/ptodata/2/pna/US6031_COMB.seq:*
65: /cgn2_6/ptodata/2/pna/US6032_COMB.seq:*
66: /cgn2_6/ptodata/2/pna/US60_MERGED_COMB.seq:*
67: /cgn2_6/ptodata/2/pna/US10_MERGED_COMB.seq:*
68: /cgn2_6/ptodata/2/pna/US09_MERGED_COMB.seq:*
69: /cgn2_6/ptodata/2/pna/US09_MERGED_COMB.seq:*
70: /cgn2_6/ptodata/2/pna/US09_MERGED_COMB.seq:*
71: /cgn2_6/ptodata/2/pna/US09_MERGED_COMB.seq:*
72: /cgn2_6/ptodata/2/pna/US08_MERGED_COMB.seq:*
73: /cgn2_6/ptodata/2/pna/US07_MERGED_COMB.seq:*
74: /cgn2_6/ptodata/2/pna/US06_MERGED_COMB.seq:*
75: /cgn2_6/ptodata/2/pna/PCT_MERGED_COMB.seq:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1745	100.0	1745	30	US-09-763-292-1
2	1739.6	99.6	1745	3	US-07-864-962C-2
3	1739.6	99.6	1745	6	US-08-238-821A-2
4	1739.6	99.6	1745	14	US-09-023-655-1059
5	1739.6	99.6	1745	31	US-09-880-107-2370
6	1739.6	99.6	1745	55	US-60-226-176-749
7	1739.6	99.6	1745	56	US-60-233-468-749
8	1739.6	99.6	1745	64	US-60-313-371-749
9	1739.6	99.6	1745	18	US-09-488-127-58
10	1739.6	99.6	1748	18	US-09-488-127-58
11	1525.2	93.1	1940	27	US-09-698-010-12366
12	1534.6	87.9	1845	14	PCT-US98-16979-4
13	1534.6	87.9	1845	15	US-09-023-655-1057
14	1534.6	87.9	1845	15	US-09-135-296-4
15	1534.6	87.9	1845	55	US-60-226-176-753
16	1534.6	87.9	1845	55	US-60-233-468-753
17	1534.6	87.9	1845	64	US-60-313-371-753
18	1534.6	87.9	1852	3	US-07-864-962C-10
19	1534.6	87.9	1852	6	US-08-238-821A-10
20	1534.6	87.9	1857	18	US-09-471-275-6
21	1529.8	87.6	1854	3	US-07-864-962C-4
22	1529.8	87.6	1854	6	US-08-238-821A-4
23	1529.8	87.6	1854	14	US-09-023-655-1056
24	1529.8	87.6	1854	31	US-09-880-107-2371
25	1512.6	86.6	1916	50	US-60-172-373-8646
26	1466.6	84.0	1473	55	US-60-226-176-750
27	1466.6	84.0	1473	55	US-60-233-468-750
28	1466.6	84.0	1473	64	US-60-313-371-750
29	1458.6	83.5	1473	6	US-08-277-031A-15
30	1458.6	83.5	1473	6	US-08-277-031B-15
31	1453.2	83.2	1835	54	US-60-213-360-5066

Db	1201	AGAAATTTCCCAACCCAGAGATGTTTGACCTCGCTCACTTTCTGGATGAAGGTGGAAATT	1260
OY	1261	TTAAGAAAAGTAACTACTTCATGCTTTCTCAGCAGAAAACGGAATTTGTTGGGAGAGG	1320
Db	1261	TTAAGAAAAGTAACTACTTCATGCTTTCTCAGCAGAAAACGGAATTTGTTGGGAGAGG	1320
OY	1321	gctggccgcgaaggagctggtttattctctgacctctcaatttacagaaactttaacctga	1380
Db	1321	GCTGGCCCGCAAGGAGCTGTTTTATTCCTTGACCTTCATTTTACAGAACTTTAACTCTGA	1380
OY	1381	aatctctgattgaccccaaggaccttgacacaactcctgtgttcaatggatttgcctctg	1440
Db	1381	AATCTCTGATTGACCCAAAGGACCTTTGACACAACCTCTGTTGTCAATGGATTGCTTCTG	1440
OY	1441	tccaccctctatcaagctgctgcctcattcctgctctgaagaagcaagaatgctggctg	1500
Db	1441	TCCGCGCCCTTCATCAGCTGCTGCTTCATTCCTGTCTGAAGAAGCACAGATGGTCTGGCTG	1500
OY	1501	ctcctgctgctccctgcagctctcttctcctgtgtccaaatttcacatactctgtgatgct	1560
Db	1501	CTCCTGTGCTGCCCTGCAGCTCTTTCCCTCTGGTCCAAATTTCACTATCTGTGATGCT	1560
OY	1561	tctctgcacgctcaatctcaatatttcccttcccccaagatctagtgaacattcagcctc	1620
Db	1561	TCCTTGACCCGTCACTCAATTTCCCTTCCCCCAAGATCTAGTGAACATTTACGCCCTC	1620
OY	1621	cattaaaaagtttcactgtgcaaatatctctgctatccccactctctaataagttac	1680
Db	1621	CATTAAAAAGTTTCACTGTGCCAAATATATCTGCTATTTCCCATCATCTATATAGTTAC	1680
OY	1681	attgagtgccacataaagtctgatacttctgcttaattgttgagctattaacatactattatta	1740
Db	1681	ATTGAGTGCCACATAAAGTCTGATACTTGTCTAAATGTTGAGTTATTAAACATAITATTATTA	1740
OY	1741	aataga	1746
Db	1741	AATAGA	1746

RESULT 3
 US-08-238-821A-2
 : Sequence 2, Application US/08238821A
 : GENERAL INFORMATION:
 : APPLICANT: GOLDSTEIN, Joyce A.
 : TITLE OF INVENTION: CLONING AND EXPRESSION OF COMPLEMENTARY
 : DNAS FOR MULTIPLE MEMBERS OF THE HUMAN CYTOCHROME P450 2C
 : TITLE OF INVENTION: SUBFAMILY
 : NUMBER OF SEQUENCES: 44
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Townsend and Townsend Khourie and Crew
 : STREET: 379 Lytton Avenue
 : CITY: Palo Alto
 : STATE: California
 : COUNTRY: US
 : ZIP: 94301
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: PatentIn Release #1.0, Version #1.25
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/238,821A
 : FILING DATE:
 : CLASSIFICATION: 435
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: US 08/201,118
 : FILING DATE: 22-FEB-1994
 : APPLICATION NUMBER: US 07/864,962
 : FILING DATE: 09-APR-1992
 : ATTORNEY/AGENT INFORMATION:
 : NAME: Liebeschuetz, Joe
 : REGISTRATION NUMBER: 37,505


```
Db 301 ATTTCACCTGGCTGAAGAGCTAACAGAGGATTTGGAAATCGTTTTTCAGCAATGGAAGA 360
Oy 361 gatgaagaggaattcgggtttctccctcatgacgtcggaatttttggatggggaaga 420
Db 361 GATGGAAGGAGATCCGGCGTTTCTCCCTCATGACGCTGGGAATTTTGGATGGGGAAGA 420
Oy 421 ggagcattgaggaccgtgttcaagagagaagccgcgtgccttggagagggttgagaaaa 480
Db 421 GGACCATTTAGGACCGTGTTCAGAGGAGCCCGCTGCCCTTGTGGAGGAGTTGAGAAAAA 480
Oy 481 ccaaggttcacccctgtagtcccaactttcatcctggggtgtgtcctcctgcaatgtagct 540
Db 481 CCAAGGCTTCACCCGTGTATCCACATTTTCATCTGGGCTGTCTCCCTGCAATGTGATCT 540
Oy 541 gctccattatttccagaacggttgcgattataaagatcacaaatttcttaacttgatgg 600
Db 541 GCTCCATTATTTCCAGAAACGTTTCGATTATAAAGATCAGCAATTTCTTAACTTGTATGG 600
Oy 601 aaaaattgaatgaataacatcagattgtagcaacccccctggatcccgatatacaataatt 660
Db 601 AAAAAATTGAATGAATAACATCAGGATTTGAAGCACCCCTGGATCCAGATATCAATAATT 660
Oy 661 ttcccactatcattgattatttcccggaacccatacaaaaattacttaaaaccttgcct 720
Db 661 TTCCCACATCATTTGATTATTTTCCCGGGAACCCATACAAATTTACTTAAAAACCTTGTCT 720
Oy 721 ttatggaagtgtatatttggagaagtaaaagacacacaaataatcgatggacatcaaca 780
Db 721 TTATGGAAGTGATATTTTGGAGAAAGTAAAGAACACCCAGATTCGATGGACATCAACA 780
Oy 781 accctcgggaacttattgattgtctctctgatcaaaatggaggaagaaagcaaaacccaac 840
Db 781 ACCCTCGGACCTTTATTGATTGTCTCTGATCAAAATGGAGAAGGAAAGCAAAACCAAC 840
Oy 841 agtctgaattcactatgtaaaacttggtaactcactgcagctgaacttacttgagctggga 900
Db 841 AGTCTGAATTCATTTGAAACTTGGTAATCCTGACGCTGACTTACTTGGAGCTGGGA 900
Oy 901 cagagacaagaacacccctgagatgtctctctctctctctctctctctctctctctctct 960
Db 901 CAGAGACAAGAGCAACCCCTGAGATATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 960
Oy 961 tcacagtaaatccaggaagagattgaactgtctctctctctctctctctctctctctctct 1020
Db 961 TCACAGCTAAAGTCCAGGAAGAGATTGAACGTTGTCATTGGCAGAAACCGAGCCCTGCA 1020
Oy 1021 tgcacagacagggggccactgcctcacacagatgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1080
Db 1021 TGCAGSACAGGGGCCACATGCCCTACACAGATGCTGTGTGTGTGTGTGTGTGTGTGTGTGT 1080
Oy 1081 tcgacctatccccaccagctgcccagtcagtgacctgtgacctgtgacctgtgacctgtgac 1140
Db 1081 TCGACTCATCTCCACAGCTTGCCTTGCCTTGCCTTGCCTTGCCTTGCCTTGCCTTGCCTTGC 1140
Oy 1141 acctcattcccaaggcccaaccattataacttccctcacttctctctctctctctctctct 1200
Db 1141 ACCTCATTTCCCAAGGGCACACCACTATTAACTTCCCTCAGCTTCTGTCTCATGACACAA 1200
Oy 1201 aagaatttcccaaccagagatttgaacctgtcacttctcacttctcacttctcacttctc 1260
Db 1201 AGAATTTCCCAACCCAGAGATTTGACCCCTGACCTGCTGCTGCTGCTGCTGCTGCTGCTG 1260
Oy 1261 ttaagaagaataactacttctatgccttctcagcagaaaaacagatttgtgtgtgtgtgt 1320
Db 1261 TTAAGAAGAATAACTTACTTCTGCTTCTCAGCAGGAACCGATTGTGTGGAGAGG 1320
Oy 1321 gcttggcccgatggagctgttttatctcgtgaccttctattttacagaactttaacctga 1380
Db 1321 GCTTGGCCGATGGAGCTGTTTTTATTCGTGACCTTCATTTTACAGAACTTTTAACTGA 1380
Oy 1381 aatctctgattgacccaagacctgacacaactctgtgtgtgtgtgtgtgtgtgtgtgtgt 1440
Db 1381 AATCTCTGATTGACCCCAAGACCTTGACACAACTCTCTGTGTGTGTGTGTGTGTGTGTGT 1440
```

RESULT 5

US-09-880-107-2370

: Sequence 2370, Application US/09880107

: GENERAL INFORMATION:

: APPLICANT: Horne, Darci T.

: APPLICANT: Vockley, Joseph G.

: APPLICANT: Scherif, Uwe

: APPLICANT: Gene Logic, Inc.

: TITLE OF INVENTION: Gene Expression Profiles In Liver Cancer

: FILE REFERENCE: 44921-5028-WO

: CURRENT APPLICATION NUMBER: US/09/880,107

: PRIOR FILING DATE: 2001-06-14

: PRIOR APPLICATION NUMBER: US 60/211,379

: PRIOR FILING DATE: 2000-06-14

: PRIOR APPLICATION NUMBER: US 60/237,054

: NUMBER OF SEQ ID NOS: 3950

: SOFTWARE: PatentIn Ver. 2.1

: SEQ ID NO 2370

: LENGTH: 1745

: TYPE: DNA

: ORGANISM: Homo sapiens

: FEATURE:

: OTHER INFORMATION: Genbank Accession No. M61854

US-09-880-107-2370

Query Match 99.5%; Score 1739.6; DB 31; Length 1746;

Best Local Similarity 99.8%; Pred. No. 0;

Matches 1742; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 1 ctccaatggaatccttttgggtccttctgtctctgtctctcatgttcttctctcttca 60

Db 1 ctccaatggaatccttttgggtccttctgtctctgtctctcatgttcttctcttca 60

Oy 61 tctggagacagagctctctgggagagaaaactcctctctggcccaactcctctccagtg 120

Db 61 tctggagacagagctctctgggagagaaaactcctctctggcccaactcctctccagtg 120

Oy 121 ttggaaatactctcacagatagatataaggaatgcagaaactccttaaccaatctctca 180

Db 121 ttggaaatactctcacagatagatataaggaatgcagaaactccttaaccaatctctca 180

Oy 181 aaatctatggccctgtgttctcactctgtattttggcctggaacgcatggtgctgcatg 240

Db 181 aaatctatggccctgtgttctcactctgtattttggcctggaacgcatggtgctgcatg 240

```
QY 241 gatagaagtggtgaaggaagccctgattgatcttggagaagagtttctggaagagcc 300
D 241 gatagaagtggtgaaggaagccctgattgatcttggagaagagtttctggaagagcc 300
QY 301 atttccactggctgaagagctaaacagaggtttggaatcgttttcaacaatgaaaga 360
D 301 atttccactggctgaagagctaaacagaggtttggaatcgttttcaacaatgaaaga 360
QY 361 gatggaagagattccggcgttttccctcatgacgctgcgggaattttgggaggggaa 420
D 361 gatggaagagattccggcgttttccctcatgacgctgcgggaattttgggaggggaa 420
QY 421 ggagcattgagaccgtgttcaagagaagaccgctgccttctggaaggagttgaaaaaa 480
D 421 ggagcattgagaccgtgttcaagagaagaccgctgccttctggaaggagttgaaaaaa 480
QY 481 ccaaggcttcacccgtgatcccaactttcatcctggcgtgctccctcgaatgagatc 540
D 481 ccaaggcttcacccgtgatcccaactttcatcctggcgtgctccctcgaatgagatc 540
QY 541 gtccattatttccagaagcgtttcgattataaaagatcacgaatttcttaacttgatg 600
D 541 gtccattatttccagaagcgtttcgattataaaagatcacgaatttcttaacttgatg 600
QY 601 aaaaattgaatgaacacatcaggattgtaagcacccttgatccagatatacaataatt 660
D 601 aaaaattgaatgaacacatcaggattgtaagcacccttgatccagatatacaataatt 660
QY 661 ttccactatcattgattatttcccggaacccataaacaatttacttaaaacccctgctt 720
D 661 ttccactatcattgattatttcccggaacccataaacaatttacttaaaacccctgctt 720
QY 721 ttatgaaagtataatttggagaagtaaaagaacaccaaagaatcgatggacataca 780
D 721 ttatgaaagtataatttggagaagtaaaagaacaccaaagaatcgatggacataca 780
QY 781 accctcggaactttattgattgcttctctgatacaaatgagaagaaagcaaaacccac 840
D 781 accctcggaactttattgattgcttctctgatacaaatgagaagaaagcaaaacccac 840
QY 841 agtctgaattcactattgaaacttggtaactcactgagctgacttacttggagctgga 900
D 841 agtctgaattcactattgaaacttggtaactcactgagctgacttacttggagctgga 900
QY 901 cagagacaacagcaacccctgagatgcttctctctctctctctctctgagcaccagag 960
D 901 cagagacaacagcaacccctgagatgcttctctctctctctctctctgagcaccagag 960
QY 961 tcacagctaaagtccaggagaagattgaaactgtcattggcagaacccggagccctgca 1020
D 961 tcacagctaaagtccaggagaagattgaaactgtcattggcagaacccggagccctgca 1020
QY 1021 tgacagacagggccacatgcctcacagatgctgtggtgacagaggtccagagataca 1080
D 1021 tgacagacagggccacatgcctcacagatgctgtggtgacagaggtccagagataca 1080
QY 1081 tgacactatcccccagcctgccccatgcaatgacgtgacgtgacgtttaaattcagaact 1140
D 1081 tgacactatcccccagcctgccccatgcaatgacgtgacgtgacgtttaaattcagaact 1140
QY 1141 acctcattcccaagggaacacattataacttccctcacttctgctacatgacaaca 1200
D 1141 acctcattcccaagggaacacattataacttccctcacttctgctacatgacaaca 1200
QY 1201 aagaatttcccaaccagagattgttgacctcgtcactttctgatatgaagtggaatt 1260
D 1201 aagaatttcccaaccagagattgttgacctcgtcactttctgatatgaagtggaatt 1260
QY 1261 ttaagaaaagtactacttcttcttctcagcaggaacacggatttggggagaggg 1320
D 1261 ttaagaaaagtactacttcttcttctcagcaggaacacggatttggggagaggg 1320
QY 1321 gctggcccgcatggagctgtttttatttcttgaccttattttacagaactttaacctga 1380
```

```
D 1321 gctggcccgcatggagctgtttttatttccctgaccttattttacagaactttaacctga 1380
QY 1381 aatctctgattgaccccaagaccttgacacaaactcctgttgcataatggattgtctctg 1440
D 1381 aatctctgattgaccccaagaccttgacacaaactcctgttgcataatggattgtctctg 1440
QY 1441 tcccgccttctatcagctgtgcttctcctctgctgaaagacacagatggtctgctg 1500
D 1441 tcccgccttctatcagctgtgcttctcctctgctgaaagacacagatggtctgctg 1500
QY 1501 ctctctgctgctccctgcagctctcttctctctggtccaaatttcaactatctgtagct 1560
D 1501 ctctctgctgctccctgcagctctcttctctctggtccaaatttcaactatctgtagct 1560
QY 1561 tcttctgaccgtcatctcacatttcccttcccccagatctagtgaacattcagctc 1620
D 1561 tcttctgaccgtcatctcacatttcccttcccccagatctagtgaacattcagctc 1620
QY 1621 cattaaaaagtcttcaactgtgcaaatatctgctattccccatactctataaagttac 1680
D 1621 cattaaaaagtcttcaactgtgcaaatatctgctattccccatactctataaagttac 1680
QY 1681 attgaatgccacataatgctgatactgtctaaatgtgagttattataattattatta 1740
D 1681 attgaatgccacataatgctgatactgtctaaatgtgagttattataattattatta 1740
QY 1741 aataga 1746
D 1741 aataga 1746
```

```
RESULT 6
US-60-226-176-749
: Sequence 749, Application US/60226176
: GENERAL INFORMATION:
: APPLICANT: Ring, Huijun Z.
: APPLICANT: Malsen, Gareth
: APPLICANT: Townley, David
: APPLICANT: Morris, MacDonald
: TITLE OF INVENTION: Single Nucleotide Polymorphisms Associated With ADME Genes
: FILE REFERENCE: GX-0013-1 P
: CURRENT APPLICATION NUMBER: US/60/226,176
: CURRENT FILING DATE: 2000-08-16
: NUMBER OF SEQ ID NOS: 2447
: SOFTWARE: PERL Program
: SEQ ID NO 749
: LENGTH: 1746
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: misc_feature
: OTHER INFORMATION: GB:HUMCYPC219
US-60-226-176-749
```

```
Query Match 99.68: Score 1739.6; DB 55; Length 1746;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1742; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
QY 1 ctccaatggatcccttttgggtccttggctctgtctctctctcatgttctctctctctctca 60
D 1 ctccaatggatcccttttgggtccttggctctgtctctctctcatgttctctctctctctca 60
QY 61 tctggagacagagctctgggagagagaaactccctcctggcccaactctctccacatga 120
D 61 tctggagacagagctctgggagagagaaactccctcctggcccaactctctccacatga 120
QY 121 ttggaataatctcacagatagatattaggatgcaacaataccttaaccaactctca 180
D 121 ttggaataatctcacagatagatattaggatgcaacaataccttaaccaactctca 180
QY 181 aaatctatggccctgtgttcaactctgtattttggcctggaaacgcatggtggtgcatg 240
```

Db 181 aatcctatggccctggttccactctgtattttggcctgaaacgcatgggtgctgcatg 240
OY 241 gatataagtggtggaagccctgattgatcttgagagagagtttcttgaaagagcc 300
Db 241 gatataagtggtggaagccctgattgatcttgagagagagtttcttgaaagagcc 300
OY 301 atttccactggctgaaagagcttaacagagagatttggatcgttttcagcaatgaaaga 360
Db 301 atttccactggctgaaagagcttaacagagagatttggatcgttttcagcaatgaaaga 360
OY 361 gatggaagagattcgctgttttccctcctatgacgctgcggaattttggatggggaaga 420
Db 361 gatggaagagattcgctgttttccctcctatgacgctgcggaattttggatggggaaga 420
OY 421 ggaagcattgagagaccgtttcaagagagaagcccgctgccttggagagagttgagaaaaa 480
Db 421 ggaagcattgagagaccgtttcaagagagaagcccgctgccttggagagagttgagaaaaa 480
OY 481 ccaagcttcacccctgattcccaactttcatctcggtgctgcctccctgcgaatgatct 540
Db 481 ccaagcttcacccctgattcccaactttcatctcggtgctgcctccctgcgaatgatct 540
OY 541 gttccatttttccagaacagtttgcattataaagatcacgaattttcttaacttgatgg 600
Db 541 gttccatttttccagaacagtttgcattataaagatcacgaattttcttaacttgatgg 600
OY 601 aaaaaattgaatgaataacacatcaggattgtaagcacccctggatccagatatgcaataatt 660
Db 601 aaaaaattgaatgaataacacatcaggattgtaagcacccctggatccagatatgcaataatt 660
OY 661 ttcccaactatattgattatttcccggaacccataaacaatttacttaaaaccttgctt 720
Db 661 ttcccaactatattgattatttcccggaacccataaacaatttacttaaaaccttgctt 720
OY 721 ttatgaaagtatatttggagaagtaaaagaacacagaagaatcgatgacatcaaca 780
Db 721 ttatgaaagtatatttggagaagtaaaagaacacagaagaatcgatgacatcaaca 780
OY 781 accctcggaactttattgattgcttccctgatacaaaatggagaagaaagcaaaaccaa 840
Db 781 accctcggaactttattgattgcttccctgatacaaaatggagaagaaagcaaaaccaa 840
OY 841 agtctgaattcactattgaaacttggtaatacactgcagctgacttacttgagactggga 900
Db 841 agtctgaattcactattgaaacttggtaatacactgcagctgacttacttgagactggga 900
OY 901 cagagacaacagcaacccctgagatatgctctctctctctctctgctgaagcaccagagg 960
Db 901 cagagacaacagcaacccctgagatatgctctctctctctctctgctgaagcaccagagg 960
OY 961 tcacagctaaagtcaggagaagattgaacgtgtcattggcagaacccggagcccttga 1020
Db 961 tcacagctaaagtcaggagaagattgaacgtgtcattggcagaacccggagcccttga 1020
OY 1021 tgcacagcaggggcccacatgcccacacagatgctgtgtgtgcacgaggtccagagataca 1080
Db 1021 tgcacagcaggggcccacatgcccacacagatgctgtgtgtgcacgaggtccagagataca 1080
OY 1081 tgcacacatcccccacagcgtgcccacatgagctgacgtgacgtttaaattcagaact 1140
Db 1081 tgcacacatcccccacagcgtgcccacatgagctgacgtgacgtttaaattcagaact 1140
OY 1141 acctatttcccaaggcacaacacatttaacttccctcacttctgtgtcatcatgacaaca 1200
Db 1141 acctatttcccaaggcacaacacatttaacttccctcacttctgtgtcatcatgacaaca 1200
OY 1201 aagaatttcccaacccagagatggttgacctcgtcacttcttctggatgaggtggaaatt 1260
Db 1201 aagaatttcccaacccagagatggttgacctcgtcacttcttctggatgaggtggaaatt 1260
OY 1261 ttaagaaagttaactacttcatgcttcttctcagcagaaacagatttgggtggagagg 1320

Db 1261 ttaagaaagttaactacttcatgcttctcagcaggagaaacaggatttgtgtggagagg 1320
OY 1321 gcttgagcccgatggagctgtttttattcctgaccttcattttacagaactttaacctga 1380
Db 1321 gcttgagcccgatggagctgtttttattcctgaccttcattttacagaactttaacctga 1380
OY 1381 aatctctgattgacccaaagacaccttgacacaaactcctgtgtgcaatggatttgcttctg 1440
Db 1381 aatctctgattgacccaaagacaccttgacacaaactcctgtgtgcaatggatttgcttctg 1440
OY 1441 tccgccccttctatcagctgcttccatctctgtctgtctgaaagacacagatggctggctg 1500
Db 1441 tccgccccttctatcagctgcttccatctctgtctgtctgaaagacacagatggctggctg 1500
OY 1501 ctctctgtcttccctcctgacgtctcttctctgtgtccaaatttcactatctatgatgct 1560
Db 1501 ctctctgtcttccctcctgacgtctcttctctgtgtccaaatttcactatctatgatgct 1560
OY 1561 tcttctgaccctgcatctcacatttcccttcccccaagatctagtgaacatlcagcctc 1620
Db 1561 tcttctgaccctgcatctcacatttcccttcccccaagatctagtgaacatlcagcctc 1620
OY 1621 cattaaaaaagtcttcaactgtgcaaatatatactgtctgtcaaatctccccatactataatagttac 1680
Db 1621 cattaaaaaagtcttcaactgtgcaaatatatactgtctgtcaaatctccccatactataatagttac 1680
OY 1681 attgagtgccacataatgctgatacttgtcttaattgttgaatttataacatattattatta 1740
Db 1681 attgagtgccacataatgctgatacttgtcttaattgttgaatttataacatattattatta 1740
OY 1741 aataga 1746
Db 1741 aataga 1746

RESULT 7
US-60-233-468-749
: Sequence 749. Application US/60233468
: GENERAL INFORMATION:
: APPLICANT: Ring, Huijun Z.
: APPLICANT: Malsen, Gareth
: APPLICANT: Townley, David
: APPLICANT: Morris, MacDonald
: APPLICANT: Valdes, Ana
: TITLE OF INVENTION: Single Nucleotide Polymorphisms Associated With ADME Genes
: FILE REFERENCE: GX-0013-2 P
: CURRENT APPLICATION NUMBER: US/60/233,468
: CURRENT FILING DATE: 2000-09-18
: NUMBER OF SEQ ID NOS: 2488
: SOFTWARE: PERL Program
: SEQ ID NO 749
: LENGTH: 1746
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: misc_feature
: OTHER INFORMATION: GB:HUMCYPC219
US-60-233-468-749

Query Match 99.6%; Score 1739.6; DB 56; Length 1746;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1742; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 cttaaatgatactcttgggtccttggctgtctgtctctgtctctcatgtttgtctctcttcaa 60
Db 1 cttaaatgatactcttgggtccttggctgtctgtctctcatgtttgtctctcttcaa 60
OY 61 tctggagacagactctggagagagaaactcctctcctggcccaactcctctccagtgga 120
Db 61 tctggagacagactctggagagagaaactcctctcctggcccaactcctctccagtgga 120
OY 121 ttggaataatcctcacagatagatttaagagatgctcagcaaaatcccttaaccaatctctcaa 180

```
Db 121 ttggaataatcctcagatagatatattaggatgctcagcaaatccttaaccaatctctcaa 180
Oy 181 aaatctatgcccctgtgttcaactctgtattttggccctggaaacgcatggtgtgctgcatg 240
Db 181 aaatctatgcccctgtgttcaactctgtattttggccctggaaacgcatggtgtgctgcatg 240
Oy 241 gatataagtggtgaagaagccctgattgatctgtgagaggagatttcttggaaagagcc 300
Db 241 gatataagtggtgaagaagccctgattgatctgtgagaggagatttcttggaaagagcc 300
Oy 301 atttcccactggctgaaagagctaaacagagagatttggaaatcgttttccagcaatggaaaga 360
Db 301 atttcccactggctgaaagagctaaacagagagatttggaaatcgttttccagcaatggaaaga 360
Oy 361 gatagaaggaatctggcgttttccctctcatgacgctgcggaaattttggatggggaaga 420
Db 361 gatagaaggaatctggcgttttccctctcatgacgctgcggaaattttggatggggaaga 420
Oy 421 ggaacattgagaccgtgttcaagaggaagaccgctgcttggagagattgaagaaaaa 480
Db 421 ggaacattgagaccgtgttcaagaggaagaccgctgcttggagagattgaagaaaaa 480
Oy 481 ccaaggcttcaacctgtgatcccaacttcatctctgggctgtgctccctgcaatgtgatct 540
Db 481 ccaaggcttcaacctgtgatcccaacttcatctctgggctgtgctccctgcaatgtgatct 540
Oy 541 gctccattatttccagaaacgttttcgattataaagatacagaaatttcttaacttgaatg 500
Db 541 gctccattatttccagaaacgttttcgattataaagatacagaaatttcttaacttgaatg 500
Oy 601 aaaaattgaatgaaacatcagattgtgaagcaccctggatccagatatgcaataatt 550
Db 601 aaaaattgaatgaaacatcagattgtgaagcaccctggatccagatatgcaataatt 550
Oy 661 ttcccactatcattgattatttccgggaacccatacaaaattactttaaacccttgctt 720
Db 661 ttcccactatcattgattatttccgggaacccatacaaaattactttaaacccttgctt 720
Oy 721 ttatggaaatgatattttggagaagatgaaagaaaccccaagaaatcgatggacatacaaa 780
Db 721 ttatggaaatgatattttggagaagatgaaagaaaccccaagaaatcgatggacatacaaa 780
Oy 781 accctcgggacttattgattgttctctgatcaaaatggagaaggaagaaagcaaaacaaac 840
Db 781 accctcgggacttattgattgttctctgatcaaaatggagaaggaagaaagcaaaacaaac 840
Oy 841 agtctgaattcactatttgaaaacttggtaatcactgcagctgacttacttggagctggga 900
Db 841 agtctgaattcactatttgaaaacttggtaatcactgcagctgacttacttggagctggga 900
Oy 901 cagagacaacaaacacacccctgagatatgctctctctctctctctctctctctctctctct 950
Db 901 cagagacaacaaacacacccctgagatatgctctctctctctctctctctctctctctctct 950
Oy 961 tcacagctaaagtccagggaagatgaaacgtgtctatggcagaacccgagccctctgca 1020
Db 961 tcacagctaaagtccagggaagatgaaacgtgtctatggcagaacccgagccctctgca 1020
Oy 1021 tgacagacagggccacatgccctcacacagatgctgtgtgtgcacgaggttccagagataca 1080
Db 1021 tgacagacagggccacatgccctcacacagatgctgtgtgtgcacgaggttccagagataca 1080
Oy 1081 tcgacctcatcccccacagcctgcccctgacgtgacctgtgacgttaaatcagaacct 1140
Db 1081 tcgacctcatcccccacagcctgcccctgacgtgacctgtgacgttaaatcagaacct 1140
Oy 1141 acctatctccaaaggccacaacacattaaacttccctcacttctgtgtgtacatgacaaca 1200
Db 1141 acctatctccaaaggccacaacacattaaacttccctcacttctgtgtgtacatgacaaca 1200
Oy 1201 aagaatttcccaaccagagattttgacccctgacacacacacacacacacacacacacacac 1260
```

```
Db 1201 aagaatttcccaaccagagatgttggaccctcgttctctctctctctctctctctctctctct 1260
Oy 1261 ttaagaaaagtaactcttcatgctcttctcagcagagaaaacgagatttctgtgtggagaggg 1320
Db 1261 ttaagaaaagtaactcttcatgctcttctcagcagagaaaacgagatttctgtgtggagaggg 1320
Oy 1321 gccggcccgcatggagctgtgtttttattctcctgacacacacacacacacacacacacacacac 1380
Db 1321 gccggcccgcatggagctgtgtttttattctcctgacacacacacacacacacacacacacacac 1380
Oy 1381 aatctctgattgaccccaagacacacacacacacacacacacacacacacacacacacacacac 1440
Db 1381 aatctctgattgaccccaagacacacacacacacacacacacacacacacacacacacacacac 1440
Oy 1441 tcccggcccttctatcagcgtgtgttctcatttctcctgtgtgaaagacacagatggtctggcctg 1500
Db 1441 tcccggcccttctatcagcgtgtgttctcatttctcctgtgtgaaagacacagatggtctggcctg 1500
Oy 1501 ctctctgctgtcctcctgacacacacacacacacacacacacacacacacacacacacacacac 1560
Db 1501 ctctctgctgtcctcctgacacacacacacacacacacacacacacacacacacacacacacac 1560
Oy 1561 tctctgacccctcattctcatttctccttcccacacacacacacacacacacacacacacacacac 1620
Db 1561 tctctgacccctcattctcatttctccttcccacacacacacacacacacacacacacacacacac 1620
Oy 1621 cattaaaaaagtctcactgtgcaaaatatactctctctctctctctctctctctctctctctctct 1680
Db 1621 cattaaaaaagtctcactgtgcaaaatatactctctctctctctctctctctctctctctctctct 1680
Oy 1681 attgagtgccacataatgctgatacttctctaatgttgagttataaacatatattatta 1740
Db 1681 attgagtgccacataatgctgatacttctctaatgttgagttataaacatatattatta 1740
Oy 1741 aataga 1746
Db 1741 aataga 1746

RESULT 8
US-60-313-371-749
: Sequence 749, Application US/60313371
: GENERAL INFORMATION:
: APPLICANT: Ring, Huijun 2.
: APPLICANT: Malsen, Gareth
: APPLICANT: Townley, David
: APPLICANT: Morris, MacDonald
: TITLE OF INVENTION: Single Nucleotide Polymorphisms Associated With ADME Genes
: FILE REFERENCE: GX-0013-5 P
: CURRENT APPLICATION NUMBER: US/60/313.371
: NUMBER OF SEQ ID NOS: 2447
: SOFTWARE: PERL Program
: SEQ ID NO 749
: LENGTH: 1746
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: misc_feature
: OTHER INFORMATION: GB:HUMCYPC219
US-60-313-371-749

Query Match 99.6%; Score 1739.6; DB 64; Length 1746;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1742; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 1 ctccaatggatccttttgggtccttctgtctctctctctctctctctctctctctctctctctca 60
Db 1 ctccaatggatccttttgggtccttctgtctctctctctctctctctctctctctctctctctca 60
Oy 61 tctgagacagagctctctggagagagaaaactcctctctgcccactcctctccagtg 120
```


Oy 61 tctggagacagactctggagaggagaaactccctctctgcccactctctctccactga 120
Db 66 tctggagacagactctggagaggagaaactccctctctgcccactctctctccactga 125
Oy 121 ttggaataatcctcacagatagattaaagatgcagcaaatccttaaccaatctctcaa 180
Db 126 ttggaataatcctcacagatagattaaagatgcagcaaatccttaaccaatctctcaa 185
Oy 181 aaatctatgcccctgtgttcaactctgtattttggcctggaaacatggtgtgctcatg 240
Db 186 aggtctatgcccctgtgttcaactctgtattttggcctggaaacatggtgtgctcatg 245
Oy 241 gatatgaatggtggaaggaagcctgattgattcttggagaggaggtttctggaagagcc 300
Db 246 gatatgaagcagtggaaggaagcctgattgattcttggagaggaggtttctggaagagca 305
Oy 301 atttcccactggctggaagagcttaacagagagatttgaatcgttttcagcaatggaaaga 360
Db 306 ttttcccactggctggaagagcttaacagagagatttgaatcgttttcagcaatggaaaga 365
Oy 361 gatggaagggatttcggcgtttctccctcatgacgctgcggaattttggatgggaaga 420
Db 366 aatggaagggatttcggcgtttctccctcatgacgctgcggaattttggatgggaaga 425
Oy 421 ggacattgaggaccgtgttcaagaggaagaccgctgcttctgtagagagttcagaaaaa 480
Db 426 ggaacattgaggaccgtgttcaagaggaagaccgctgcttctgtagagagttcagaaaaa 485
Oy 481 ccaaggttccacctgtgacccacttctccctctggtggtgctccctgcgaatgcatct 540
Db 486 ccaaggttccacctgtgacccacttctccctctggtggtgctccctgcgaatgcatct 545
Oy 541 gctccatttttccagaaagtttccatataaagatcacgaattcttaacttgaatg 600
Db 546 gctccatttttccagaaagtttccatataaagatcacgaattcttaacttgaatg 605
Oy 601 aaaaattgaataaaacatcaggatttgaagaccctctggtatccagatgcgaataatt 660
Db 606 aaaaattgaataaaacatcaggatttgaagaccctctggtatccagatgcgaataatt 665
Oy 661 ttccactatcattgattttccgggaacccatacaaatcttaaaactctgctt 720
Db 666 ttctccctatcattgattttccgggaacccatacaaatcttaaaactctgctt 725
Oy 721 ttatggaagtgatttttggagaagtaaaagaaaccccaagaaatcagatgcacatca 780
Db 726 ttatggaagtgatttttggagaagtaaaagaaaccccaagaaatcagatgcacatca 785
Oy 781 accctcggactttattgattgcttctctgatcaaaatggagaagaaacaaacacaa 840
Db 786 accctcaggactttattgattgcttctctgatcaaaatggagaagaaacaaacacaa 845
Oy 841 agtctgaattcactattgaaaacttggtaatacactgcagctgacttacttggagctgga 900
Db 846 catctgaatttactattgaaaacttggtaatacactgcagctgacttacttggagctgga 905
Oy 901 cagagacaacagacacccctgagatgctctctctctctctctcactgaacacccagag 960
Db 906 cagagacacagacacacccctgagatgctctctctctctctctcactgaacacccagag 965
Oy 961 tcacagcttaaaagtcagagaagattgaacgtgtcatttggcagaaacccggagccctgca 1020
Db 966 tcacagcttaaaagtcagagaagattgaacgtgtcatttggcagaaacccggagccctgca 1025
Oy 1021 tcacagacaggggccaatgcctcacagagatgctgtgtagtcacagagttccagagata 1080
Db 1026 tcgaagacagagggccacatgcctcacagagatgctgtgtgtcagagagttccagagata 1085
Oy 1081 tcgacctcatcccaccagctgcccctgagtgacctgtgacctgttaattcagaaact 1140
Db 1086 ttgaccttctcccaccagctgcccctgagtgacctgtgacctgttaattcagaaact 1145
Oy 1141 acctcatttcccaggggcaaacacattataacttccctcacttctgtgctacatgaaca 1200

Db 1146 atctcatttcccaggggcaaacattataatttccctgactctgtgctacatgaaca 1205
Oy 1201 aagaatttcccacccagagatgtttgacctcgtcactttcttgatgaagtggaact 1260
Db 1206 aagaatttcccacccagagatgtttgacctcgtcactttcttgatgaagtggaact 1265
Oy 1261 ttaaaaaagtaactactactcattcgtcttctcagcaggaaaacggatttgtgtggagagg 1320
Db 1266 ttaaaaaagtaactactactcgtcttctcagcaggaaaacggatttgtgtggagagg 1325
Oy 1321 gctcggcccgatgagctgttttattctcctgaccttattttacagaactttaacctga 1380
Db 1326 cctcggcccgatgagctgttttattctcctgaccttattttacagaactttaacctga 1385
Oy 1381 aatctctgattgacccaaaggaccttgacaaactcctctgttcaalgaattgtctctg 1440
Db 1386 aatctctgattgacccaaaggaccttgacaaactcctctgttcaalgaattgtctctg 1445
Oy 1441 tcccggccctctatcagctgtcttctcattctctgtctgaagaagacagatggtctgctg 1500
Db 1446 tcccggccctctatcagctgtcttctcattctctgtctgaagaagacagatggtctgctg 1505
Oy 1501 ctctctgctctcctcagctctcttctctctggtggtggtggtggtggtggtggtggtg 1553
Db 1506 ctgctgtgagctcctcagctctcttctctctggtggtggtggtggtggtggtggtg 1565
Oy 1554 tgaagcttctcagcctcagctcattcacaatttcccttccccaagatctagtgaaact 1613
Db 1566 taatgcttctcagcctcagctcattcacaatttcccttccccaagatctagtgaaact 1625
Oy 1614 cagctccacta-----aaaaagtttcaactgtcgaataatactatctatcccca 1663
Db 1626 cagctccactaagagagaggttctctgttctcactgtgcaaatatactatctatccca 1685
Oy 1664 tactctataatgattcagtgacacataatgctgatactgtcttaatttgagttga 1723
Db 1686 tactctgtaacagttcagtgactgacacataatgctatactatactatctatggttga 1745
Oy 1724 ttaacattattattataaataga 1746
Db 1746 ttaattgttattattataaataga 1768

RESULT 13

US-09-023-655-1057

: Sequence 1057, Application US/09023655

: GENERAL INFORMATION:

: APPLICANT: Cocks, Benjamin G.

: APPLICANT: Susan G. Stuart

: APPLICANT: Jeffrey J. Seilhamer

: TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF BLOOD CELL GENE

: TITLE OF INVENTION: EXPRESSION

: NUMBER OF SEQUENCES: 1508

: CORRESPONDENCE ADDRESS:

: ADDRESSEE: INCYTE PHARMACEUTICALS, INC.

: STREET: 3174 PORTER DRIVE

: CITY: PALO ALTO

: STATE: CALIFORNIA

: COUNTRY: USA

: ZIP: 94304

: COMPUTER READABLE FORM:

: MEDIUM TYPE: Floppy disk

: COMPUTER: IBM PC compatible

: OPERATING SYSTEM: PC-DOS/MS-DOS

: SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2

: CURRENT APPLICATION DATA: US/09/023, 655

: FILING DATE: HEREWITH

: CLASSIFICATION:

: PRIOR APPLICATION DATA:

: APPLICATION NUMBER:

: FILING DATE:


```

US-09-135-296-4
: Sequence 4, Application US/09135296
: GENERAL INFORMATION:
: APPLICANT: Jones, Jeffrey P.
: APPLICANT: Shimoji, Miyuki
: TITLE OF INVENTION: FUNCTIONAL BACTERIAL/MAMMALIAN CYTOCHROME P450 CHIMERA
: FILE REFERENCE: 176/60231
: CURRENT APPLICATION NUMBER: US/09/135,296
: CURRENT FILING DATE: 1998-08-17
: EARLIER APPLICATION NUMBER: 60/056,754
: EARLIER FILING DATE: 1997-08-20
: NUMBER OF SEQ ID NOS: 14
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 4
: LENGTH: 1845
: TYPE: DNA
: ORGANISM: mammalian
US-09-135-296-4

Query Match      87.9%; Score 1534.6; DB 15; Length 1845;
Best Local Similarity 92.9%; Pred. No. 0;
Matches 163; Conservative 0; Mismatches 109; Indels 17; Gaps

Qy 1 cttcaatggatccttttggctctgtgctctgtctctctcatgttctgtctctcttca 50
Db 6 cttcaatggatccttttggctctgtgctctgtctctctcatgttctgtctcttca 65

Qy 61 tctggagacagactctgggagggaaaaactccctctctggcccaactctctccagtga 120
Db 66 tctggagacagactctgggagggaaaaactccctctctggcccaactctctccagtga 125

Qy 121 ttggaaatcctcacagatagattaaaggatgtcagcaaatccttaaccaatctctca 180
Db 126 ttggaaatcctcacagatagattaaaggacatcagcaaatccttaaccaatctctca 185

Qy 181 aaatctatggccctgtgttcaactctgtattttggcctggaacgcatggtgctgcagt 240
Db 186 aggtctatggccctgtgttcaactctgtattttggcctgaaaccatagtggtgctgcagt 245

Qy 241 gatataagtgggtgaaggaaagccctgattgtatcttggagaggagatttctggaaaggcc 300
Db 246 gatataagcagtggaaggaaagccctgattgtatcttggagaggagatttctggaaaggca 305

Qy 301 atttcccaactgctgaagagagctaacagaggatttggaaatcggttccagcaatggaaaga 360
Db 306 ttttcccaactgctgaagagagctaacagaggatttggaaatgttttcagcaatggaaaga 365

Qy 361 gatggaagagatctggcgctttctccctcatgacgctgcggaatttggagatgggaaaga 420
Db 366 aatggaagagatccggcgctttctccctcatgacgctgcggaatttggagatgggaaaga 425

Qy 421 gaagcatggagaccgttgttcaagagaagcccgctccttggaggagttgagaaaaa 480
Db 426 gaagcatggagaccgttgttcaagagaagcccgctccttggaggagttgagaaaaa 485

Qy 481 ccaaggttcaacctgtgatcccaacttcatctggagctgtgctccctgcaatgtgatct 540
Db 486 ccaaggttcaacctgtgatcccaacttcatctggagctgtgctccctgcaatgtgatct 545

Qy 541 gttccattatttccagaagaactgttcgatataaagatcacgaatttcttaactgtatgg 600
Db 546 gttccattatttccagaagaactgttcgatataaagatcacgaatttcttaactgtatgg 605

Qy 601 aaaaattgaatgaacaatcaggattgtgaagacccttgatccagatatacataatt 660
Db 606 aaaaattgaatgaacaatcaggattgtgaagacccttgatccagatatacataatt 665

Qy 661 ttcccaactatcatgtattttccggggaaccatacaaaaattacttaaaaaccttgct 720
Db 666 ttctctcatcatctgatttctccgggaactcaaaaaattacttaaaaaccttgct 725

Qy 721 ttatggaaagtgtattttggagaagtaaaagaacacccaagaatctgtggacatcaaca 780

```

US-60-226-176-753
: Sequence 753, Application US/60226176
: GENERAL INFORMATION:
: APPLICANT: Ring, Huijun Z.
: APPLICANT: Malsen, Gareth
: APPLICANT: Townley, David
: APPLICANT: Morris, MacDonald
: TITLE OF INVENTION: Single Nucleotide Polymorphisms Associated With ADME Genes
: FILE REFERENCE: GX-0013-1 P
: CURRENT APPLICATION NUMBER: US/60/226,176
: CURRENT FILING DATE: 2000-08-16
: NUMBER OF SEQ ID NOS: 2447
: SOFTWARE: PERL Program
: SEQ ID NO 753
: LENGTH: 1845
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: misc.feature
: OTHER INFORMATION: GB:HUMCYP2C9A
US-60-226-176-753

Query Match 87.9%; Score 1534.6; DB 55; Length 1845;
Best Local Similarity 92.9%; Pred. No. 0;
Matches 1637; Conservative 0; Mismatches 109; Indels 1; Gaps 2;

```
Qy 1 cttcaatgagctcttattgctctgtctgtctgtctctcatgtttgtctctctctcaaa 50
Db 6 ctccaatgagctcttattgctctgtctgtctctctcatgtttgtctctctctcaaa 55
Qy 61 tctgagacagagctctgagagagaaactccctccctgcccactctctcccaatga 120
Db 66 tctgagacagagctctgagagagaaactccctccctgcccactctctcccaatga 125
Qy 121 ttggaatatctacagatagatttaaggatgctcagcaaatcttcaaccactctcaaa 180
Db 126 ttggaatatctacagatagatttaaggatgctcagcaaatcttcaaccactctcaaa 185
Qy 181 aaatcatgcccctgttccactctgtattttggctggaacgcatgtgtgtctgcatg 240
Db 186 aggtctatggcccctgttccactctgtattttggctggaacgcatgtgtgtctgcatg 245
Qy 241 gatagaagtggtgaagaagccctgattgatcttggagagagatcttctggaagagccc 300
Db 246 gatagaagcagtggaagaagccctgattgatcttggagagagatcttctggaagagcca 305
Qy 301 atttcccactgctgaagagactaacagagattttggaatcgttttcagcaaatggaaga 360
Db 306 ttttcccactgctgaagagactaacagagattttggaatggttttcagcaaatggaaga 365
Qy 361 gatgaagagatctgcgcttctccctcatgacgctgcgaattttggagatgggaaga 420
Db 366 aatggaagagatctgcgcttctccctcatgacgctgcggaattttggagatgggaaga 425
Qy 421 ggagcatgagagaccgtgttcaagaggaagcccgcctgctgtgagagattgagaaaaa 480
Db 426 ggagcatgagagaccgtgttcaagaggaagcccgcctgctgtgagagattgagaaaaa 485
Qy 481 ccaagcttccactctgaatccacttcatctcctgagctgtactccctgcaatgcatct 540
Db 486 ccaagcttccactctgaatccacttcatctcctgagctgtactccctgcaatgcatct 545
Qy 541 gctccattatttccagaaagcttgcgattataaagatcacgaaatttcttaactttgagg 500
Db 546 gctccattatttccagaaagcttgcgattataaagatcacgaaatttcttaactttgagg 505
Qy 601 aaaaattgaataaaacatcagagattgttaagcaccctcctgcatccagatagcaaat 560
Db 606 aaaaattgaataaaacatcagagattgttgaagcagccctgcatccagatagcaaat 565
Qy 661 ttcccactatcatgattatttcccgggaagcccataaacaatttacttaaaccttact 720
Db 1746 ttaatatgttattattataaataga 1768
```

```
Db 566 ttctctctatcatgtattacttcccgggaactcacacaaattacttaaaaaacttgctt 725
Qy 721 ttatggaagtgatatttttgagaaagttaaagaacacccaagaatcgtgacatcaaca 780
Db 726 ttatggaagtgatatttttgagaaagttaaagaacacccaagaatcgtgacatgaaca 785
Qy 781 accctcggagctttattgattgctctctctgatacaaaatggagaagaaagcaaacacac 840
Db 786 accctcggagctttattgattgctctctctgatacaaaatggagaagaaagcaaacacac 845
Qy 841 agtctgaattcactattgaaaacttgtaactcactgcagctgacttacttggagctggga 900
Db 846 catctgaatttactattgaaagcttggaaaacactgcagttgacttgtttggagctggga 905
Qy 901 cagagacaacaagcacaaacctgagatatgctctctctctctctctctdaagcaccacaggg 960
Db 906 cagagacgacaagcacaaacctgagatatgctctctctctctctctctdaagcaccacaggg 965
Qy 961 tcacagctaaagtccaggaagaagattgaacgtgctcattggcagaaaacggagcccttgca 1020
Db 966 tcacagctaaagtccaggaagaagattgaacgtgctcattggcagaaaacggagcccttgca 1025
Qy 1021 tgcacgacaggggcccacatgcccctacacagatgctgtggtgcacgaggtccagagataca 1080
Db 1026 tgcacgacaggggcccacatgcccctacacagatgctgtggtgcacgaggtccagagataca 1085
Qy 1081 tgcacctctatccccacacagcctgcccctgacgtgcagctgacgtttaaattcagaacct 1140
Db 1086 tgcacctctctccccacacagcctgcccctgacgtgcagctgacgtttaaattcagaacct 1145
Qy 1141 acctcattccccaaagggcacaaccattataacttccctcactctctgtctacatgacaaca 1200
Db 1146 acctcattccccaaagggcacaaccattataacttccctcactctctgtctacatgacaaca 1205
Qy 1201 aagaatttcccaaccccagagatgtttgacctcgtcactttcttgatgaaggtggaaatt 1260
Db 1206 aagaatttcccaaccccagagatgtttgacctcactcactttcttgatgaaggtggcaatt 1265
Qy 1261 ttaagaaagttaactacttcatgctcttctcagcagaaaacaggatttgtgtggagagg 1320
Db 1266 ttaagaaagttaactacttcatgctcttctcagcagaaaacaggatttgtgtggagagg 1325
Qy 1321 gcttggccctcatgagctgttttattcctgcacttcatctttacagaaccttaacctga 1380
Db 1326 ccttggccctcatgagctgttttattcctgcacttcatctttacagaaccttaacctga 1385
Qy 1381 aatctctgattgaccccaaaagaccttgacacaaactcctgtgttcaatggattgctctg 1440
Db 1386 aatctctgattgaccccaaaagaccttgacacaaactcctgtgttcaatggattgctctg 1445
Qy 1441 tcccgccttctatacagctgtgcttcatctctgtctgaagaacacagatggctggctg 1500
Db 1446 tggcgcccttctatacagctgtgcttcatctctgtctgaagaagagagagctggctggctg 1505
Qy 1501 ctctctgtgctcctgcagctctcttctctctgg-----tccaaatttcaatatctg 1553
Db 1506 ctgctgtgcatcctcgcagctctcttctctctggcattatccactcttctcaatatctg 1565
Qy 1554 tgaactcttctgaaccctcatctcacttttcccttcccacagatcagagaacct 1613
Db 1556 taatgcttcttctcactctcatctcacttttcccttcccacagatcagagaacct 1625
Qy 1614 cagctccatt-----aaaaagtttctgctgcaaatatctctatctcccca 1663
Db 1626 cgaactccattacggagagtttctctgttctcactgttgcataatctctctattctcca 1685
Qy 1664 tactctataatgattacattgaagtcacataaattgctgatacttctgcttaatttgagttta 1723
Db 1686 tactctgtaacagttgcatgtgactgtcacataaattgctcactattctatcttattgagttta 1745
Qy 1724 ttaacataattattataaataga 1746
Db 1746 ttaatatgttattattataaataga 1768
```

Search completed: April 19, 2002, 11:02:26
Job time: 13246 sec

Result No.	Score	Query Match	Length	DB	ID	Description	
C 1	40.4	5.4	13227	6	US-10-105-299-7201	Sequence 7201, App	
C 2	35	4.7	11145	6	US-10-105-299-12104	Sequence 12104, A	
C 3	34.6	4.7	3220	6	US-10-103-313-100	Sequence 100, App	
C 4	33.8	4.5	266	5	US-09-975-254-20507	Sequence 20507, A	
C 5	33.2	4.5	52845	6	US-10-105-299-12317	Sequence 12317, A	
C 6	33	4.4	34762	6	US-10-105-299-11849	Sequence 11849, A	
C 7	32.8	4.4	272	6	US-10-102-524-480	Sequence 480, App	
C 8	32.2	4.3	261	5	US-09-975-254-24327	Sequence 24327, A	
C 9	31.8	4.3	68571	7	US-60-368-184-1	Sequence 1, Appli	
C 10	31.6	4.3	555	6	US-10-106-698-786	Sequence 786, App	
C 11	31	4.2	266	5	US-09-789-189-1604	Sequence 1604, Ap	
C 12	31	4.2	6833	6	US-10-105-299-7023	Sequence 7023, Ap	
C 13	30.8	4.1	658	6	US-10-102-805-359	Sequence 359, App	
C 14	30.8	4.1	748	5	US-10-106-698-1347	Sequence 1347, Ap	
C 15	30.6	4.1	49375	6	US-10-105-299-12428	Sequence 12428, A	
C 16	30.4	4.1	671	5	US-10-106-698-1786	Sequence 1786, Ap	
C 17	29.8	4.0	1134	6	US-10-105-299-1744	Sequence 1744, Ap	
C 18	29.8	4.0	2007	6	US-10-106-698-1481	Sequence 1481, Ap	
C 19	29.8	4.0	2893	5	US-10-105-299-11321	Sequence 11321, A	
C 20	29.8	4.0	31304	6	US-10-105-299-12035	Sequence 12035, A	
C 21	29.6	4.0	271	5	US-09-975-254-11869	Sequence 11869, A	
C 22	29.6	4.0	51961	6	US-10-105-299-6419	Sequence 6419, Ap	
C 23	29.4	4.0	1306	6	US-10-105-299-14391	Sequence 14391, A	
C 24	29.4	4.0	1325	6	US-10-105-299-2832	Sequence 2832, Ap	
C 25	29.4	4.0	2037	5	US-09-978-403A-423	Sequence 423, App	
C 26	29.4	4.0	2037	5	US-09-978-444A-429	Sequence 429, App	


```

Query: Match      4.3%   Score 31.8;   DB 7;   Length 49571;
Best Local Similarity: 57.0%;   Pred. No. 27;
Matches 77;   Conservative 0;   Mismatches 57;   Indels 1;   Gaps 1;

QY 555 gttcaaggaggagccgcgtgcctctgtggaggagttgagaaaaccaaagatgggtgacc 525
    || ||||| ||||| || ||||| ||||| || ||||| ||||| || ||||| |||||
DB 1191 GTGAAGAGCAGAGCTGTGGATCTCTGGAGCAATGTTATAGCAGAGGAGAGAGTGCC 1132

QY 526 tactcatatcac-tgaacttactgaactactatcttcttctactgacatcttggaaaca 584
    || || || || || || || || || || || || || || || || || || || || ||
DB 1131 AAGGCCCTGCACAGAGGGGCTTCCTGCCTGGTGTATTCCAGGAATAGCTTCAGGGGAACA 1072

QY 495 ttccaaggagtgacca 599
    || || || || || || || || || || || || || || || || || || || || ||

```


GenCore version 4.5
Copyright (C) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 19, 2002, 11:02:26 ; Search time 5830.82 Seconds
(without alignments)
2760.419 Million cell updates/sec

Title: US-09-763-292-2

Perfect score: 743

Sequence: 1 tcagaaatattgaagcctg.....ttagctcatgtgaagcgggg 743

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 22023303 seqs, 10831430700 residues

Total number of hits satisfying chosen parameters: 44046606

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending Patents NA_Main:*

1: /cgn2_6/ptodata/2/pna/US0611_COMB.seq:*

2: /cgn2_6/ptodata/2/pna/US06_COMB.seq:*

3: /cgn2_6/ptodata/2/pna/US07_COMB.seq:*

4: /cgn2_6/ptodata/2/pna/US080_COMB.seq:*

5: /cgn2_6/ptodata/2/pna/US081_COMB.seq:*

6: /cgn2_6/ptodata/2/pna/US082_COMB.seq:*

7: /cgn2_6/ptodata/2/pna/US083_COMB.seq:*

8: /cgn2_6/ptodata/2/pna/US084_COMB.seq:*

9: /cgn2_6/ptodata/2/pna/US085_COMB.seq:*

10: /cgn2_6/ptodata/2/pna/US086_COMB.seq:*

11: /cgn2_6/ptodata/2/pna/US087_COMB.seq:*

12: /cgn2_6/ptodata/2/pna/US088_COMB.seq:*

13: /cgn2_6/ptodata/2/pna/US089_COMB.seq:*

14: /cgn2_6/ptodata/2/pna/US090_COMB.seq:*

15: /cgn2_6/ptodata/2/pna/US091_COMB.seq:*

16: /cgn2_6/ptodata/2/pna/US092_COMB.seq:*

17: /cgn2_6/ptodata/2/pna/US093_COMB.seq:*

18: /cgn2_6/ptodata/2/pna/US094_COMB.seq:*

19: /cgn2_6/ptodata/2/pna/US095A_COMB.seq:*

20: /cgn2_6/ptodata/2/pna/US095B_COMB.seq:*

21: /cgn2_6/ptodata/2/pna/US095C_COMB.seq:*

22: /cgn2_6/ptodata/2/pna/US095D_COMB.seq:*

23: /cgn2_6/ptodata/2/pna/US096A_COMB.seq:*

24: /cgn2_6/ptodata/2/pna/US096B_COMB.seq:*

25: /cgn2_6/ptodata/2/pna/US096C_COMB.seq:*

26: /cgn2_6/ptodata/2/pna/US096D_COMB.seq:*

27: /cgn2_6/ptodata/2/pna/US096E_COMB.seq:*

28: /cgn2_6/ptodata/2/pna/US097A_COMB.seq:*

29: /cgn2_6/ptodata/2/pna/US097B_COMB.seq:*

30: /cgn2_6/ptodata/2/pna/US097C_COMB.seq:*

31: /cgn2_6/ptodata/2/pna/US098_COMB.seq:*

32: /cgn2_6/ptodata/2/pna/US099_COMB.seq:*

33: /cgn2_6/ptodata/2/pna/US6000_COMB.seq:*

34: /cgn2_6/ptodata/2/pna/US6001_COMB.seq:*

35: /cgn2_6/ptodata/2/pna/US6002_COMB.seq:*

36: /cgn2_6/ptodata/2/pna/US6003_COMB.seq:*

37: /cgn2_6/ptodata/2/pna/US6004_COMB.seq:*

38: /cgn2_6/ptodata/2/pna/US6005_COMB.seq:*

39: /cgn2_6/ptodata/2/pna/US6006_COMB.seq:*

40: /cgn2_6/ptodata/2/pna/US6007_COMB.seq:*

41: /cgn2_6/ptodata/2/pna/US6008_COMB.seq:*

42: /cgn2_6/ptodata/2/pna/US6009_COMB.seq:*

43: /cgn2_6/ptodata/2/pna/US6010_COMB.seq:*

44: /cgn2_6/ptodata/2/pna/US6011_COMB.seq:*

45: /cgn2_6/ptodata/2/pna/US6012_COMB.seq:*

46: /cgn2_6/ptodata/2/pna/US6013_COMB.seq:*

47: /cgn2_6/ptodata/2/pna/US6014_COMB.seq:*

48: /cgn2_6/ptodata/2/pna/US6015_COMB.seq:*

49: /cgn2_6/ptodata/2/pna/US6016_COMB.seq:*

50: /cgn2_6/ptodata/2/pna/US6017_COMB.seq:*

51: /cgn2_6/ptodata/2/pna/US6018_COMB.seq:*

52: /cgn2_6/ptodata/2/pna/US6019_COMB.seq:*

53: /cgn2_6/ptodata/2/pna/US6020_COMB.seq:*

54: /cgn2_6/ptodata/2/pna/US6021_COMB.seq:*

55: /cgn2_6/ptodata/2/pna/US6022_COMB.seq:*

56: /cgn2_6/ptodata/2/pna/US6023_COMB.seq:*

57: /cgn2_6/ptodata/2/pna/US6024_COMB.seq:*

58: /cgn2_6/ptodata/2/pna/US6025_COMB.seq:*

59: /cgn2_6/ptodata/2/pna/US6026_COMB.seq:*

60: /cgn2_6/ptodata/2/pna/US6027_COMB.seq:*

61: /cgn2_6/ptodata/2/pna/US6028_COMB.seq:*

62: /cgn2_6/ptodata/2/pna/US6029_COMB.seq:*

63: /cgn2_6/ptodata/2/pna/US6030_COMB.seq:*

64: /cgn2_6/ptodata/2/pna/US6031_COMB.seq:*

65: /cgn2_6/ptodata/2/pna/US6032_COMB.seq:*

66: /cgn2_6/ptodata/2/pna/US60_MERGED_COMB.seq:*

67: /cgn2_6/ptodata/2/pna/US10_MERGED_COMB.seq:*

68: /cgn2_6/ptodata/2/pna/US09_MERGED_COMB.seq:*

69: /cgn2_6/ptodata/2/pna/US09_MERGED_COMB.seq:*

70: /cgn2_6/ptodata/2/pna/US09_MERGED_COMB.seq:*

71: /cgn2_6/ptodata/2/pna/US09_MERGED_COMB.seq:*

72: /cgn2_6/ptodata/2/pna/US08_MERGED_COMB.seq:*

73: /cgn2_6/ptodata/2/pna/US07_MERGED_COMB.seq:*

74: /cgn2_6/ptodata/2/pna/US06_MERGED_COMB.seq:*

75: /cgn2_6/ptodata/2/pna/US06_MERGED_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	743	100.0	743	30	US-09-763-292-2	Sequence 2, Appli
2	739.8	99.6	743	55	US-60-226-176-756	Sequence 756, App
3	739.8	99.6	743	56	US-60-233-458-756	Sequence 756, App
4	739.8	99.6	743	64	US-60-313-371-756	Sequence 15, Appl
5	728.8	98.1	7609	54	US-60-212-657-15	Sequence 56, Appl
6	728.8	98.1	32768	54	US-60-213-181-56	Sequence 24, Appl
c 7	708	95.3	51955	57	US-60-248-498-24	Sequence 61, Appl
c 8	708	95.3	222537	57	US-60-248-542-62	Sequence 62, Appl
c 9	708	95.3	222537	57	US-60-245-227-9	Sequence 9, Appl
c 10	708	95.3	225532	57	US-09-488-127-59	Sequence 59, Appl
11	622.6	83.8	8437	18	US-09-488-127-59	Sequence 59, Appl
12	622.6	83.8	8437	18	US-09-488-127B-59	Sequence 97, Appl
13	622.6	83.8	8437	18	US-09-488-127B-97	Sequence 147, App
14	555.8	74.8	572	51	US-60-182-895-147	Sequence 148, App
15	555.8	74.8	572	51	US-60-182-895-148	Sequence 149, App
16	555.8	74.8	572	51	US-60-182-895-150	Sequence 150, App
17	555.8	74.8	572	51	US-60-182-895-151	Sequence 151, App
18	555.8	74.8	572	51	US-60-182-895-152	Sequence 152, App
19	555.8	74.8	572	51	US-60-198-818-18	Sequence 18, Appl
20	555.8	74.8	572	51	US-60-213-795-31	Sequence 31, Appl
21	544.2	73.2	1580	54	US-60-213-795-31	Sequence 46, Appl
22	544.2	73.2	1580	54	US-60-195-048-108	Sequence 108, App
c 23	532.4	71.7	722	52	US-60-195-048-109	Sequence 109, App
c 24	532.4	71.7	722	52	US-60-205-169-3	Sequence 3, Appl
c 25	521.2	70.1	710	53	US-60-198-818-178	Sequence 178, App
26	488.2	65.7	664	52	US-09-671-317-49	Sequence 49, Appl
27	433.6	58.4	1001	26	US-60-229-511-5	Sequence 5, Appl
c 28	433.6	58.4	32141	55	US-60-248-498-23	Sequence 23, Appl
29	433.6	58.4	36191	57	US-60-245-227-10	Sequence 10, Appl
30	433.6	58.4	36584	57	US-60-248-498-59	Sequence 59, Appl
31	433.6	58.4	118047	57	US-60-248-498-59	

```

32 433.6 58.4 118047 57 US-60-248-542-60
33 421.8 56.8 1001 26 US-09-671-317-50
34 415.4 55.9 436 26 US-09-671-317-55
c 35 369 49.7 734 18 US-09-488-127-64
c 36 369 49.7 734 18 US-09-488-127B-64
37 366.4 49.3 605 53 US-60-205-169-6
38 366.2 49.3 557 18 US-09-488-127-62
39 366.2 49.3 557 18 US-09-488-127B-62
c 40 346.4 46.6 16238 54 US-60-212-657-14
41 331.2 44.6 654 55 US-60-226-176-740
42 331.2 44.6 654 56 US-60-233-468-740
43 331.2 44.6 654 64 US-60-313-371-740
c 44 325 43.7 1595 53 US-60-200-392-43
c 45 325 43.7 1595 53 US-60-200-392-44

```

ALIGNMENTS

```

RESULT 1
US-09-763-292-2
: Sequence 2, Application US/09763292
: GENERAL INFORMATION:
: APPLICANT: AB Sangtec Medical
: TITLE OF INVENTION: New method
: FILE REFERENCE: primers
: CURRENT APPLICATION NUMBER: US/09/763.292
: CURRENT FILING DATE: 2000-02-21
: NUMBER OF SEQ ID NOS: 19
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 2
: LENGTH: 743
: TYPE: DNA
: ORGANISM: Homo sapiens
US-09-763-292-2

```

```

Query Match 100.0%; Score 743; DB 30; Length 743;
Best Local Similarity 100.0%; Pred. No. 3.7e-210;
Matches 743; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tcagaaatattgaagcctgtgtgctgaataaaagcacaatacaatacaataatca 60
Db 1 tcagaaatattgaagcctgtgtgctgaataaaagcacaatacaatacaataatca 60

Qy 61 tgcataatcaggcttagcaaatgacaaaatagtaacttcgtttgttctgtctgtct 120
Db 61 tgcataatcaggcttagcaaatgacaaaatagtaacttcgtttgttctgtctgtct 120

Qy 121 acttcttagctctcaaggctctatggccctgtgttcaactctgtattttggcctgaacc 180
Db 121 acttcttagctctcaaggctctatggccctgtgttcaactctgtattttggcctgaacc 180

Qy 181 catagtgtctcatgatgatgaagcagtggaaggaagccctgattgatcttggagaga 240
Db 181 catagtgtctcatgatgatgaagcagtggaaggaagccctgattgatcttggagaga 240

Qy 241 gtttctggaagaggcattttcccaactggctgaaagagctaaacagagagatttgtagtg 300
Db 241 gtttctggaagaggcattttcccaactggctgaaagagctaaacagagagatttgtagtg 300

Qy 301 tgcattgtctgttccagcattctcttgggagatggggagatggaaaaacagagacttac 360
Db 301 tgcattgtctgttccagcattctcttgggagatggggagatggaaaaacagagacttac 360

Qy 361 agagctctcgggcagagcttggccaccacatggctgccagtgccagcttccctttt 420
Db 361 agagctctcgggcagagcttggccaccacatggctgccagtgccagcttccctttt 420

Qy 421 ctgctcgggagatctccctccttagtttcttctctctctggttaggaattgttttcagcaa 480
Db 421 ctgctcgggagatctccctccttagtttcttctctctctggttaggaattgttttcagcaa 480

```

```

Oy 481 tggaaagaaatggaagagagatcccgctttctccctcatgacgctgcggaatttgggat 540
Db 481 tggaaagaaatggaagagagatcccgctttctccctcatgacgctgcggaatttgggat 540

Qy 541 gggaaagagagcattgaggaaccgtgttcaagagaagagcccgctgtcttggaggagtt 600
Db 541 gggaaagagagcattgaggaaccgtgttcaagagaagagcccgctgtcttggaggagtt 600

Qy 601 gagaaaccaaaggtgtgggtgaacctactccatcatcatcactgaccttactgactactatct 660
Db 601 gagaaaccaaaggtgtgggtgaacctactccatcatcatcactgaccttactgactactatct 660

Qy 661 tctctactgacattcttgaaacatttcagggtggccatattcttccattatgagctctg 720
Db 661 tctctactgacattcttgaaacatttcagggtggccatattcttccattatgagctctg 720

Qy 721 ttgttagctcatgtgaagcgggg 743
Db 721 ttgttagctcatgtgaagcgggg 743

```

RESULT 2

```

US-60-226-176-756
: Sequence 756, Application US/60226176
: GENERAL INFORMATION:
: APPLICANT: Ring, Huijun 2.
: APPLICANT: Malsen, Gareth
: APPLICANT: Townley, David
: APPLICANT: Morris, MacDonald
: TITLE OF INVENTION: Single Nucleotide Polymorphisms Associated With ADME Genes
: FILE REFERENCE: GX-0013-1 P
: CURRENT APPLICATION NUMBER: US/60/226,176
: CURRENT FILING DATE: 2000-08-16
: NUMBER OF SEQ ID NOS: 2447
: SOFTWARE: PERL Program
: SEQ ID NO 756
: LENGTH: 743
: TYPE: DNA
: ORGANISM: Homo sapiens
: NAME/KEY: misc.feature
: OTHER INFORMATION: GB:HUM2C9X02
US-60-226-176-756

```

Query Match 99.6%; Score 739.8; DB 55; Length 743;
Best Local Similarity 99.7%; Pred. No. 3.3e-209;
Matches 741; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

Qy 1 tcagaaatattgaagcctgtgtgctgaataaaagcacaatacaatacaataatca 60
Db 1 tcagaaatattgaagcctgtgtgctgaataaaagcacaatacaatacaataatca 60

Qy 61 tgcataatcaggcttagcaaatgacaaaatagtaacttcgtttgttctgtctgtct 120
Db 61 tgcataatcaggcttagcaaatgacaaaatagtaacttcgtttgttctgtctgtct 120

Qy 121 acttcttagctctcaaggctctatggccctgtgttcaactctgtattttggcctgaacc 180
Db 121 acttcttagctctcaaggctctatggccctgtgttcaactctgtattttggcctgaacc 180

Qy 181 catagtgtctcatgatgatgaagcagtggaaggaagccctgattgatcttggagaga 240
Db 181 catagtgtctcatgatgatgaagcagtggaaggaagccctgattgatcttggagaga 240

Qy 241 gtttctggaagaggcattttcccaactggctgaaagagctaaacagagagatttggtagtg 300
Db 241 gtttctggaagaggcattttcccaactggctgaaagagctaaacagagagatttggtagtg 300

Qy 301 tgcattgtctgttccagcattctcttgggagatggggagatggaaaaacagagacttac 360
Db 301 tgcattgtctgttccagcattctcttgggagatggggagatggaaaaacagagacttac 360

```

```
Oy 361 agagctctcggcagagcttgcccatccacatgctgcccagtgctcagcttctctttt 420
Db 361 agagctctcggcagagcttgcccatccacatgctgcccagtgctcagcttctcttt 420
Oy 421 ctgctcggatctccctcctcagtttctctctctctctctctctctctctctct 480
Db 421 ctgctcggatctccctcctcagtttctctctctctctctctctctctctctct 480
Oy 481 tggaaagaatggaagagatcccgctttctccctcctcctcctcctcctcctcct 540
Db 481 tggaaagaatggaagagatcccgctttctccctcctcctcctcctcctcctcct 540
Oy 541 ggggaagaggagcattgaggaccgtgttcaagagaaagcccgctgctctgtggagagt 600
Db 541 ggggaagaggagcattgaggaccgtgttcaagagaaagcccgctgctctgtggagagt 600
Oy 601 gaaaaaaacaaaggtggtgacctactcctacatcctcactgaccttactgacctact 660
Db 601 gaaaaaaacaaaggtggtgacctactcctacatcctcactgaccttactgacctact 660
Oy 661 tctctactgacattcttggaaacatttcagggggtggccatctcttctattagctctg 720
Db 661 tctctactgacattcttggaaacatttcagggggtggccatctcttctattagctctg 720
Oy 721 ttgttagctcatgtgaagcgggg 743
Db 721 ttgttagctcatgtgaagcgggg 743
```

```
RESULT 3
US-60-233-468-756
; Sequence 756, Application US/60233468
; GENERAL INFORMATION:
; APPLICANT: Ring, Huijun Z.
; APPLICANT: Malsen, Gareth
; APPLICANT: Townley, David
; APPLICANT: Morris, MacDonald
; APPLICANT: Valdes, Ana
; TITLE OF INVENTION: Single Nucleotide Polymorphisms Associated With ADME Genes
; FILE REFERENCE: GX-0013-2 P
; CURRENT APPLICATION NUMBER: US/60/233,468
; CURRENT FILING DATE: 2000-09-18
; NUMBER OF SEQ ID NOS: 2488
; SOFTWARE: PERL Program
; SEQ ID NO 756
; LENGTH: 743
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: GB:HUM2C9X02
US-60-233-468-756
```

```
Query Match 99.6%; Score 739.8; DB 56; Length 743;
Best Local Similarity 99.7%; Pred. No. 3.3e-209;
Matches 741; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 tcagaaatattgaagcctgtggtgataaaagcacaatacaatacaatacaataatca 60
Db 1 tcagaaatattgaagcctgtggtgataaaagcacaatacaatacaatacaataatca 60
Oy 61 tgcataatcaggcttagcaaatggaacaaatagtaactctgttctgttctctgtct 120
Db 61 tgcataatcaggcttagcaaatggaacaaatagtaactctgttctgttctctgtct 120
Oy 121 acttctcagctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 180
Db 121 acttctcagctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 180
Oy 181 catagtgtgctcagatgaagcagtggaagcctcctcctcctcctcctcctcctcct 240
Db 181 catagtgtgctcagatgaagcagtggaagcctcctcctcctcctcctcctcctcct 240
```

```
Oy 241 gtttctggaagagcattttccacatggctgaaagagcacaagagatttggtaggtg 300
Db 241 gtttctggaagagcattttccacatggctgaaagagcacaagagatttggtaggtg 300
Oy 301 tgcattgctcctggttcagcattctgtcttggggatggggagagatggaaaacagagacttac 360
Db 301 tgcattgctcctggttcagcattctgtcttggggatggggagagatggaaaacagagacttac 360
Oy 361 agagctcctcgggcagagcttgcccatccacatggctgcccagtgctcagcttctcttt 420
Db 361 agagctcctcgggcagagcttgcccatccacatggctgcccagtgctcagcttctcttt 420
Oy 421 ctgctcggatctccctcctcagtttctctctctctctctctctctctctctctctct 480
Db 421 ctgctcggatctccctcctcagtttctctctctctctctctctctctctctctctct 480
Oy 481 tggaaaagaatggaagagatcccgctttctccctcctcctcctcctcctcctcctcct 540
Db 481 tggaaaagaatggaagagatcccgctttctccctcctcctcctcctcctcctcctcct 540
Oy 541 ggggaagaggagcattgaggaccgtgttcaagagaaagcccgctgctctgtggagagt 600
Db 541 ggggaagaggagcattgaggaccgtgttcaagagaaagcccgctgctctgtggagagt 600
Oy 601 gaaaaaaacaaaggtggtgacctactcctacatcctcactgaccttactgacctactat 660
Db 601 gaaaaaaacaaaggtggtgacctactcctacatcctcactgaccttactgacctactat 660
Oy 661 tctctactgacattcttggaaacatttcagggggtggccatctcttctattagctctg 720
Db 661 tctctactgacattcttggaaacatttcagggggtggccatctcttctattagctctg 720
Oy 721 ttgttagctcatgtgaagcgggg 743
Db 721 ttgttagctcatgtgaagcgggg 743
```

```
RESULT 4
US-60-313-371-756
; Sequence 756, Application US/60313371
; GENERAL INFORMATION:
; APPLICANT: Ring, Huijun Z.
; APPLICANT: Malsen, Gareth
; APPLICANT: Townley, David
; APPLICANT: Morris, MacDonald
; TITLE OF INVENTION: Single Nucleotide Polymorphisms Associated With ADME Genes
; FILE REFERENCE: GX-0013-5 P
; CURRENT APPLICATION NUMBER: US/60/313,371
; CURRENT FILING DATE: 2001-08-16
; NUMBER OF SEQ ID NOS: 2447
; SOFTWARE: PERL Program
; SEQ ID NO 756
; LENGTH: 743
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: GB:HUM2C9X02
US-60-313-371-756
```

```
Query Match 99.6%; Score 739.8; DB 64; Length 743;
Best Local Similarity 99.7%; Pred. No. 3.3e-209;
Matches 741; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 tcagaaatattgaagcctgtggtgataaaagcacaatacaatacaatacaataatca 60
Db 1 tcagaaatattgaagcctgtggtgataaaagcacaatacaatacaatacaataatca 60
Oy 61 tgcataatcaggcttagcaaatggaacaaatagtaactctgttctgttctctgtct 120
Db 61 tgcataatcaggcttagcaaatggaacaaatagtaactctgttctgttctctgtct 120
```



```
QY 121 acttctcagctcctcaaaaggtctatgcccctgtgttccactctgtatttttggcctgaacc 180
Db 121 acttctcagctcctcaaaaggtctatgcccctgtgttccactctgtatttttggcctgaacc 180
QY 181 catagtgtgtctgcatgatatgaagcagtgaaagcctgattgtatcttggaggaga 240
Db 181 catagtgtgtctgcatgatatgaagcagtgaaagcctgattgtatcttggaggaga 240
QY 241 gttttctgaaagagcattttccactggtgaaagagctaaacagagatttggtagtg 300
Db 241 gttttctgaaagagcattttccactggtgaaagagctaaacagagatttggtagtg 300
QY 301 tgcattgtcctgtttcagcatctgtcttgggagtggtgaaagagctaaacagagattt 360
Db 301 tgcattgtcctgtttcagcatctgtcttgggagtggtgaaagagctaaacagagattt 360
QY 361 agagctcctcggcagagctgtgcccattccacatgctgcccagtgctcagcttctctt 420
Db 361 agagctcctcggcagagctgtgcccattccacatgctgcccagtgctcagcttctctt 420
QY 421 cttgctgggattcctcctcctagtttctgttctctcttcttcttcttcttcttctt 480
Db 421 cttgctgggattcctcctcctagtttctgttctctcttcttcttcttcttcttctt 480
QY 481 tggaaagaaatggaagagatcccgcttctcctcctcctcctcctcctcctcctcct 540
Db 481 tggaaagaaatggaagagatcccgcttctcctcctcctcctcctcctcctcctcct 540
QY 541 ggggaagaggagcattgagaccgtgttcaagagagagcccgctgcttggaggagtt 600
Db 541 ggggaagaggagcattgagaccgtgttcaagagagagcccgctgcttggaggagtt 600
QY 601 gagaaaccagaggtggtgacccctactcctcctcctcctcctcctcctcctcctcct 660
Db 601 gagaaaccagaggtggtgacccctactcctcctcctcctcctcctcctcctcctcct 660
QY 661 tcttactgacattcttggaaacatttcagggtggccatattcttcttcttcttctt 720
Db 661 tcttactgacattcttggaaacatttcagggtggccatattcttcttcttcttctt 720
QY 721 ttgttagctcatgtgaagcgggg 743
Db 721 ttgttagctcatgtgaagcgggg 743

RESULT 5
US-60-212-657-15
: Sequence 15, Application US/60212657
: GENERAL INFORMATION:
: APPLICANT: Beasley, Ellen
: TITLE OF INVENTION: ISOLATED HUMAN PHASE I DRUG-METABOLIZING
: TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN
: TITLE OF INVENTION: DRUG-METABOLIZING PHASE I PROTEINS, AND USES THEREOF
: FILE REFERENCE: CL000672
: CURRENT APPLICATION NUMBER: US/60/212,657
: CURRENT FILING DATE: 2000-06-19
: NUMBER OF SEQ ID NOS: 303
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 15
: LENGTH: 7609
: TYPE: DNA
: ORGANISM: HUMAN
US-60-212-657-15

Query Match 98.1%; Score 728.8; DB 54; Length 7609;
Best Local Similarity 99.6%; Pred. No. 2e-205;
Matches 741; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 1 tcgaaataattgaagcctgtgtgctgaataaagacatacaatacaatacaataatca 50
Db 3936 tcgaaataattgaagcctgtgtgctgaataaagacatacaatacaatacaataatca 3995
```

```
QY 61 tgctaaatcaggtctagcaaatgacaaaatagtaacttcgttgcgttatctctgct 120
Db 3996 tgctaaatcaggtctagcaaatgacaaaatagtaacttcgttgcgttatctctgct 4055
QY 121 acttctcagctcctcaaaagctctatgcccctgtgttccactctgtattttggcctgaacc 180
Db 4056 acttctcagctcctcaaaagctctatgcccctgtgttccactctgtattttggcctgaacc 4115
QY 181 catagtgtgtctgcatgatatgaagcagtgaaagagccctgattgattcttggaggaga 240
Db 4116 catagtgtgtctgcatgatatgaagcagtgaaagagccctgattgattcttggaggaga 4175
QY 241 gtttctggaagagcattttccactggtgaaagagctaaacagagatttggtagtg 300
Db 4176 gtttctggaagagcattttccactggtgaaagagctaaacagagatttggtagtg 4235
QY 301 tgcattgtcctgtttcagcatctgtcttgggagtggtgaaagagctaaacagagattt 360
Db 4236 tgcattgtcctgtttcagcatctgtcttgggagtggtgaaagagctaaacagagattt 4295
QY 361 agagctcctcggcagagcttggcccattccacatgctgcccagtgctcagcttctctt 420
Db 4296 agagctcctcggcagagcttggcccattccacatgctgcccagtgctcagcttctctt 4355
QY 421 cttgctgggattcctcctcctcctagtttcttcttcttcttcttcttcttcttctt 480
Db 4356 cttgctgggattcctcctcctcctagtttcttcttcttcttcttcttcttcttctt 4415
QY 481 tggaaagaaatggaagagatcccgcttctcctcctcctcctcctcctcctcctcct 540
Db 4416 tggaaagaaatggaagagatcccgcttctcctcctcctcctcctcctcctcctcct 4475
QY 541 ggggaagaggagcattgagaccgtgttcaagagagagcccgctgcttggaggagtt 600
Db 4476 ggggaagaggagcattgagaccgtgttcaagagagagcccgctgcttggaggagtt 4535
QY 601 gagaaaccagaggtggtgacccctactcctcctcctcctcctcctcctcctcctcct 660
Db 4536 gagaaaccagaggtggtgacccctactcctcctcctcctcctcctcctcctcctcct 4595
QY 661 tcttactgacattcttggaaacatttcagggtggccatattcttcttcttcttctt 719
Db 4596 tcttactgacattcttggaaacatttcagggtggccatattcttcttcttcttctt 4655
QY 720 gttgttagctcatgtgaagcgggg 743
Db 4656 gttgttagctcatgtgaagcgggg 4679
```

```
RESULT 6
US-60-213-181-56
: Sequence 56, Application US/60213181
: GENERAL INFORMATION:
: APPLICANT: Beasley, Ellen
: TITLE OF INVENTION: ISOLATED HUMAN PHASE I DRUG-METABOLIZING
: TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN PHASE I
: TITLE OF INVENTION: DRUG-METABOLIZING PHASE I PROTEINS, AND USES THEREOF
: FILE REFERENCE: CL000690
: CURRENT APPLICATION NUMBER: US/60/213,181
: CURRENT FILING DATE: 2000-06-22
: NUMBER OF SEQ ID NOS: 261
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 56
: LENGTH: 32768
: TYPE: DNA
: ORGANISM: Human
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (1)...(32768)
: OTHER INFORMATION: n = A,T,C or G
US-60-213-181-56
```

Query Match 98.1%; Score 728.8; DB 54; Length 32768;
Best Local Similarity 99.6%; Pred. No. 4e-205;
Matches 741; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 1 tcagaaatatttgaagcctgtgtgctgaataaaagcacaatacaatacaatgaaatataca 60
DB 29095 tcagaaatatttgaagcctgtgtgctgaataaaagcacaatacaatacaatgaaatataca 29154

QY 61 tgctaaatcagccttagcaaatgacaaaatagtaactctgttgcgtgttactctgtct 120
DB 29155 tgctaaatcagccttagcaaatgacaaaatagtaactctgttgcgtgttactctgtct 29214

QY 121 acttctcagctctcaaaagctctatggccctgtgttcaactctgttatttggccctgaaacc 180
DB 29215 acttctcagctctcaaaagctctatggccctgtgttcaactctgttatttggccctgaaacc 29274

QY 181 catagtgtgtcgtcgtatgatgaagcagtgaaagagccctgattgatcttggaggaga 240
DB 29275 catagtgtgtcgtcgtatgatgaagcagtgaaagagccctgattgatcttggaggaga 29334

QY 241 gtttctggaagagcattttcccaactggctgaaagagctaaacagagagatttggtaggtg 300
DB 29335 gtttctggaagagcattttcccaactggctgaaagagctaaacagagagatttggtaggtg 29394

QY 301 tgcattgtcctgtttcagcatctgtcttgggagtgaggagagtgaaagagctaaacagagagattt 360
DB 29395 tgcattgtcctgtttcagcatctgtcttgggagtgaggagagtgaaagagctaaacagagagattt 29454

QY 361 agagctcctcggcagagcattttcccaactggctgaaagagctaaacagagagattt 420
DB 29454 agagctcctcggcagagcattttcccaactggctgaaagagctaaacagagagattt 29514

QY 421 ctgctcgggagctccctcctcagctgttcttgcctcctcctcctcctcctcctcctcctt 480
DB 29515 ctgctcgggagctccctcctcagctgttcttgcctcctcctcctcctcctcctcctcctt 29574

QY 481 tggaaagaaatggagagatccggcgttctcctcctcctcctcctcctcctcctcctcct 540
DB 29575 tggaaagaaatggagagatccggcgttctcctcctcctcctcctcctcctcctcctcct 29634

QY 541 ggggaagagagcatttgaagaccgtgttcaagagagcccgctgccttgcgtgagagatt 600
DB 29635 ggggaagagagcatttgaagaccgtgttcaagagagcccgctgccttgcgtgagagatt 29694

QY 601 gagaaacacaaaggtgggtgacctactcctacatcactgaccttactgacctactatct 560
DB 29695 gagaaacacaaaggtgggtgacctactcctacatcactgaccttactgacctactatct 29754

QY 661 tctcactgacattcttggaaacatttccaggggtgcccatatcttcttcttcttcttcttct 719
DB 29755 tctcactgacattcttggaaacatttccaggggtgcccatatcttcttcttcttcttcttct 29814

QY 720 gttgttagctcatgtgaagcgagg 743
DB 29815 gttgttagctcatgtgaagcgagg 29838

RESULT 7

US-60-248-498-61/c

; Sequence 24, Application US/60248498

; GENERAL INFORMATION:

; APPLICANT: Beasley, Ellen

; TITLE OF INVENTION: ISOLATED HUMAN PHASE I DRUG-METABOLIZING

; TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN

; TITLE OF INVENTION: DRUG-METABOLIZING PHASE I PROTEINS, AND USES THEREOF

; FILE REFERENCE: CL000910

; CURRENT APPLICATION NUMBER: US/60/248,498

; NUMBER OF SEQ ID NOS: 264

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 24

; LENGTH: 51955

; TYPE: DNA
; ORGANISM: HUMAN
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(51955)
; OTHER INFORMATION: n = A,T,C or G
US-60-248-498-24

Query Match 95.3%; Score 708; DB 57; Length 51955;
Best Local Similarity 97.8%; Pred. No. 8e-199;
Matches 728; Conservative 0; Mismatches 15; Indels 1; Gaps 1;

QY 1 tcagaaatatttgaagcctgtgtgctgaataaaagcacaatacaatacaatgaaatataca 60
DB 46910 TCAGAAATATTGAAGCCTGTGTGGCTGAATAAAGCATACAATAAATGAAATATCA 46851

QY 61 tgctaaatcagccttagcaaatgacaaaatagtaactctgttgcgtgttactctgtct 120
DB 46950 TGCTAAATCAGGCTTAGCAAAATGGACAAATAGTAACCTCGTTTGTCTGTCTCTGTCT 46791

QY 121 acttctcagctctcaaaagctctatggccctgtgttcaactctgttatttggccctgaaacc 180
DB 46790 ACTTCTCAGCTCTCAAAAGGCTATGGCCCTGTGTTCACCTCTGTATTGTGGCCTGAAACC 46731

QY 181 catagtgtgtcgtcgtatgatgaagcagtgaaagagccctgattgatcttggaggaga 240
DB 46730 CATAGTGGTGTGCATGGATATGAAGCAGTGAAGGAGCCCTGATTGATCTTGGAGAGCA 46671

QY 241 gtttctggaagagcattttcccaactggctgaaagagctaaacagagagatttggtaggtg 300
DB 46670 GTTTCTGGAAGAGGCATTTTCCCACTGGCTGAAAGAGCTAAACAGAGGATTTGGTAGGTG 46611

QY 301 tgcattgtcctgtttcagcatctgtcttgggagtgaggagagtgaaagagctaaacagagagattt 360
DB 46610 TGCATGTGCTGTTTCAGCATCTGTCTTGGGATGGGAGGATGAAACACAGAGCTTAC 46551

QY 361 agagctcctcggcagagcattttcccaactcctcctcctcctcctcctcctcctcctt 420
DB 46550 AGAGCTTCTCGGCAGAGCTTTGCCCATCCACATGGCTGCCAGAGTGTGAGCTTCTCTTT 46491

QY 421 ctgctcgggagctccctcctcagctgttcttctcctccttcttcttcttcttcttctcagcaa 480
DB 46490 CTGTGCTGGGATCTCCCTCTCTAGTTTCTGTTTCTCTCTGTAGGAATGTTTTCAGCAA 46431

QY 481 tggaaagaaatggagagatccggcgttctcctcctcctcctcctcctcctccttggagatt 540
DB 46430 TGAAGAAATGGAAGGAGATCCGGCGTTTCTCCCTCATGACGCTGCGGAATTTTGGGAT 46371

QY 541 ggggaagagagcatttgaagaccgtgttcaagagagaccccgctgccttgcgtgagagatt 600
DB 46370 GGGGAAGAGAGCATTGAGGACCGTGTTCAGAGAGAGCCCGCTGCTTGTGGAGAGATT 46311

QY 601 gagaaacacaaaggtgggtgacctactcctacatcactgaccttactgacctactatct 660
DB 46310 GAGAAACACCAAGGTGGGTGAACATACTCTCTATCACTGACCTTCTCTGGACTGCTCTCC 46251

QY 661 tctcactgacattcttggaaacatttccaggggtgcccatatcttcttcttcttcttcttct 719
DB 46250 TCTCTACTGACATCTTGGAAACATTTTCAGGGGTGCCAGATCTTTATTATTGGAGTCTCTG 46191

QY 720 gttgttagctcatgtgaagcgagg 743
DB 46190 GTTGTAGCTCATGTGAAGCAGGG 46167

RESULT 8

US-60-248-498-61/c

; Sequence 61, Application US/60248498

; GENERAL INFORMATION:

; APPLICANT: Beasley, Ellen

; TITLE OF INVENTION: ISOLATED HUMAN PHASE I DRUG-METABOLIZING

; TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN

;
; TITLE OF INVENTION: DRUG-METABOLIZING PHASE I PROTEINS, AND USES THEREOF
; FILE REFERENCE: CL000910
; CURRENT APPLICATION NUMBER: US/60/248,498
; CURRENT FILING DATE: 2000-11-15
; NUMBER OF SEQ ID NOS: 264
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 61
; LENGTH: 222537
; TYPE: DNA
; ORGANISM: HUMAN
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(222537)
; OTHER INFORMATION: n = A,T,C or G
US-60-248-498-61

Query Match 95.3%; Score 708; DB 57; Length 222537;
Best Local Similarity 97.8%; Pred. No. 1.6e-198;
Matches 728; Conservative 0; Mismatches 15; Indels 1; Gaps 1;

QY 1 tcagaaatattgaagcctgtgctgaataaaagcacaatacaatacaataatca 60
Db 90372 TCAGAAATATTGAAGCCTGTGCTGATATAAAGCATACAAATGAAATATCA 90313

QY 61 tgcataatcaggcttagcaaaatagtaacttctgttctgttctctgtct 120
Db 90312 TGCTAAATCAGGCTTAGCAAAATGACAAATAGTAACCTTGTGTTATCTGTCT 90253

QY 121 acttcttagctctaaagctctatgcccctgtgttccactgttatttggcctgaacc 180
Db 90252 ACTTCTTAGCTCTAAAGCTCTATGGCCCTGTGTTCACTCTGTATTTTGGCCTGAACC 90193

QY 181 catagtggtgctgcatggtatgaagcagtggaagcctgattgattccttgaggaga 240
Db 90192 CATAGTGGTCTGCATGGATATGAACAGTGAAGGAGCCCTGATTGATCTTGGAGAGA 90133

QY 241 gtcttctgaagaggcatcttccactggtctgaagagctaacagagatttgtagtg 300
Db 90132 GTTCTTCTGGAAGAGGCGATTTCCCACTGGCTGAAAGAGCTAACAGAGGATTTGGTAGGTG 90073

QY 301 tgcattgctctgttcagcatctgtcttgggagtgaggagtgaaacagagacttac 360
Db 90072 TGCATGTGCTGTTTCAGCATCTGCTTGGGATGGGAGGATGAACACAGAGACTTAC 90013

QY 361 agagctctcggcagagcttggcccatccacatggctgcccagtgctcagcttctctt 420
Db 90012 AGAGCTCTCTGGGACAGAGTTGGCCCATCCACATGGCTGCCAGTGTCTCTCTTT 89953

QY 421 ctgctctgggagctcctcctcctagttcgttctctctctttagaattgtttcagcaa 480
Db 89952 CTGCTCTGGGATCTCCTCTCTAGTTCTGTTCTCTCTCTTCTGTTAGGAATGTTTTCAGCAA 89893

QY 481 tggaaagaaatggaagagatccgctgttctcctcatgacgctgcggaattttgggat 540
Db 89892 TGGAAAGAAATGGAAGAGATCCGGGTTTCTCCCTCATGAGCTGCGGAATTTTGGGAT 89833

QY 541 ggggaagagagcattgagacgctgttcaagagaagacccctcctctgtgaggaatt 600
Db 89832 GGGGAAGAGAGCATTGAGACCGCTGTTCAAGAGGAGCCCGCTGCCCTTGTGGAGGATTT 89773

QY 601 gagaaacacacaggggtgggtgacctactccatatacactgaccttactggactatct 660
Db 89772 GAGAAACACAGGGGTGGGTGGGTGAACATACTCTATCACTGACCTTCTTGGACTGCTCTCC 89713

QY 661 tctctactgacattctgaaacatttcagggttgagccatatttcttacttactatgaat-ctg 719
Db 89712 TCTCTACTGACATCTTGAACATCTCAGGGGTGGCCAGATCTTTTATTTGGAGTCTCTG 89653

QY 720 gttgttagctcatgtgaagcgggg 743
Db 89652 GTTGTTAGCTCATGTGAAGCAGGG 89529

RESULT 9

US-60-248-542-62/c
; Sequence 62, Application US/60248542
; GENERAL INFORMATION:
; APPLICANT: Beasley, Ellen
; TITLE OF INVENTION: ISOLATED HUMAN PHASE I DRUG-METABOLIZING
; TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN PHASE I
; TITLE OF INVENTION: DRUG-METABOLIZING PROTEINS, AND USES THEREOF
; FILE REFERENCE: CL000946
; CURRENT APPLICATION NUMBER: US/60/248,542
; CURRENT FILING DATE: 2000-11-16
; NUMBER OF SEQ ID NOS: 342
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 62
; LENGTH: 222537
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(222537)
; OTHER INFORMATION: n = A,T,C or G
US-60-248-542-62

Query Match 95.3%; Score 708; DB 57; Length 222537;
Best Local Similarity 97.8%; Pred. No. 1.6e-198;
Matches 728; Conservative 0; Mismatches 15; Indels 1; Gaps 1;

QY 1 tcagaaatattgaagcctgtgctgaataaaagcacaatacaatacaataatca 60
Db 90372 TCAGAAATATTGAAGCCTGTGCTGATATAAAGCATACAAATGAAATATCA 90313

QY 61 tgcataatcaggcttagcaaaatagtaacttctgttctgttctctgtct 120
Db 90312 TGCTAAATCAGGCTTAGCAAAATGACAAATAGTAACCTTGTGTTATCTGTCT 90253

QY 121 acttcttagctctaaagctctatgcccctgtgttccactgttatttggcctgaacc 180
Db 90252 ACTTCTTAGCTCTAAAGCTCTATGGCCCTGTGTTCACTCTGTATTTTGGCCTGAACC 90193

QY 181 catagtggtgctgcatggtatgaagcagtggaagcctgattgattccttgaggaga 240
Db 90192 CATAGTGGTCTGCATGGATATGAACAGTGAAGGAGCCCTGATTGATCTTGGAGAGA 90133

QY 241 gtcttctgaagaggcatcttccactggtctgaagagctaacagagatttgtagtg 300
Db 90132 GTTCTTCTGGAAGAGGCGATTTCCCACTGGCTGAAAGAGCTAACAGAGGATTTGGTAGGTG 90073

QY 301 tgcattgctctgttcagcatctgtcttgggagtgaggagtgaaacagagacttac 360
Db 90072 TGCATGTGCTGTTTCAGCATCTGCTTGGGATGGGAGGATGAACACAGAGACTTAC 90013

QY 361 agagctctcggcagagcttggcccatccacatggctgcccagtgctcagcttctctt 420
Db 90012 AGAGCTCTCTGGGACAGAGTTGGCCCATCCACATGGCTGCCAGTGTCTCTCTTT 89953

QY 421 ctgctctgggagctcctcctcctagttcgttctctctttagaattgtttcagcaa 480
Db 89952 CTGCTCTGGGATCTCCTCTCTAGTTCTGTTCTCTCTCTTCTGTTAGGAATGTTTTCAGCAA 89893

QY 481 tggaaagaaatggaagagatccgctgttctcctcatgacgctgcggaattttgggat 540
Db 89892 TGGAAAGAAATGGAAGAGATCCGGGTTTCTCCCTCATGAGCTGCGGAATTTTGGGAT 89833

QY 541 ggggaagagagcattgagacgctgttcaagagaagacccctcctctgtgaggaatt 600
Db 89832 GGGGAAGAGAGCATTGAGACCGCTGTTCAAGAGGAGCCCGCTGCCCTTGTGGAGGATTT 89773

QY 601 gagaaacacacaggggtgggtgacctactccatatacactgaccttactggactatct 660
Db 89772 GAGAAACACAGGGGTGGGTGGGTGAACATACTCTATCACTGACCTTCTTGGACTGCTCTCC 89713

```
Oy 661 tcttactgacattcttgaagaacatttccagggtgagccatattcttctcattatgaat-ctg 719
Db 89712 TCTCTACTGACATCTCTGGAACATTTTCAGGGGTGCCAGATCTTTTATTTCGAGTCTTG 89653

Oy 720 gttgttagctcatgtgaagcggg 743
Db 89652 GTTGTAGCTCATGTGAAGCAGG 89629

RESULT 10
US-60-245-227-9/c
: Sequence 9, Application US/60245227
: GENERAL INFORMATION:
: APPLICANT: Beasley, Ellen
: TITLE OF INVENTION: ISOLATED HUMAN PHASE I DRUG-METABOLIZING
: TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN
: TITLE OF INVENTION: DRUG-METABOLIZING PHASE I PROTEINS, AND USES THEREOF
: FILE REFERENCE: CL000876
: CURRENT APPLICATION NUMBER: US/60/245,227
: CURRENT FILING DATE: 2000-11-03
: NUMBER OF SEQ ID NOS: 129
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 9
: LENGTH: 225532
: TYPE: DNA
: ORGANISM: HUMAN
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (1)...(225532)
: OTHER INFORMATION: n = A,T,C or G
US-60-245-227-9

Query Match 95.3%; Score 708; DB 57; Length 225532;
Best Local Similarity 97.8%; Pred. No. 1.6e-198;
Matches 728; Conservative 0; Mismatches 15; Indels 1; Gaps 1;

Oy 1 tcagaaatatttgaagcctgtgtggtgaataaaagcatacaatacaatacaataca 60
Db 93367 TCAGAAATATTTGAAGCCTGTGTGGCTGAATAAAGCATACAATAAATCAATCA 93308

Oy 61 tcttaacacagcttagcaaatgagcaaaatagtaactctgttctgttctctgtct 120
Db 93307 TCCTAATCAGCTTAGCAAAATGAGCAAAATAGTAACCTTCGTTGCTGTTAICTGTCT 93248

Oy 121 acttctcagctctcaaggtctatgcccctgtgttcaactctgtatcttctgtct 180
Db 93247 ACTTCTCAGCTCTCAAGGCTATGGCCCTGTGTCACTCTGTATTTTGGCCCTGAAC 93188

Oy 181 catagtgtgtctgcatgatgaagcagtgaggaagccctgattgtatcttgaggaga 240
Db 93187 CATAGTGTGTGTCATGATGAAGCAGTGAAGGAGCCCTGATTGATCTTGGAGAGGA 93128

Oy 241 gttttctggaagagcattttccactggctgaaagagcctaacagagatttgtagtg 300
Db 93127 GTTTTCTGGAAGAGCATTTTCCCACTGGCTGAAGAGCTAACAGAGGATTTGGTAGGTG 93068

Oy 301 tcatgtcctgtttcagcatctgtcttgggagtgaggagatgaaacacagagacttac 360
Db 93067 TGCATGTGCTGTTTTCAGCATCTGTCTTGGGATGGGAGGATGGAACACAGAGACTTAC 93008

Oy 361 agagctcctcgggagagcttggcccatccacatggctggccagtgccagcttctcttt 420
Db 93007 AGAGCTCTCGGGCAGAGCTTGGCCCATCCACATGGCTGCCAGTGTCCAGCTTCTCTTT 92948

Oy 421 ctgtcctgggagatcctcctcagttcttctctctctctgttagaattgttttcagcaa 480
Db 92947 CTTGCCCTGGGATCTCCCTCCTAGTTTCGTTTCTTCTCTCTCTCTCTCTCTCTCTCTCT 92888

Oy 481 tggaaagaaatggaagagatccgcgtttctccctcatgacgctgcggaattttgggat 540
Db 92887 TGGAAAGAAATGGAAGAGATCCGCGCTTCTCCCTCATGACCTTCGGGAAATTTGGGAT 92828
```

```
Oy 541 ggggaagagagcattgaggaccgtgttcaagagaagcccgctgttggaggagtt 600
Db 92827 GGGGAAGAGGAGCATTGAGGACCGTGTTCAGAGGAAGCCGCTGCTTGTGGAGGATT 92768

Oy 601 gagaaaaaaccaaggggtgggtgacctactccatatacactgacctgagactactatct 660
Db 92767 GAGAAAAACCAAGGGGTGGGTGAACATACTCTCTATCACTGACCTTTCTTGGAGACTCTCC 92708

Oy 661 tctctactgacattcttggaaaacatttcagggtgagccatattcttccattatgaat-ctg 719
Db 92707 TCTCTACTGACATCTTGGAAACATTTTCAGGGGTGCCAGATCTTTTATTTCGAGTCTCTG 92648

Oy 720 gttgttagctcatgtgaagcggg 743
Db 92647 GTTGTAGCTCATGTGAAGCAGG 92624

RESULT 11
US-09-488-127-59
: Sequence 59, Application US/09488127
: GENERAL INFORMATION:
: APPLICANT: Thomann, Hans-Ulrich
: APPLICANT: FitzGerald, Michael
: TITLE OF INVENTION: RAPID DETERMINATION OF GENE STRUCTURE
: TITLE OF INVENTION: USING CDNA SEQUENCE
: FILE REFERENCE: 2709.1005-000
: CURRENT APPLICATION NUMBER: US/09/488,127
: CURRENT FILING DATE: 2000-01-20
: NUMBER OF SEQ ID NOS: 96
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 59
: LENGTH: 8437
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (1)...(8437)
: OTHER INFORMATION: n = A,T,C or G
US-09-488-127-59

Query Match 83.8%; Score 622.6; DB 18; Length 8437;
Best Local Similarity 93.3%; Pred. No. 9.7e-174;
Matches 694; Conservative 0; Mismatches 45; Indels 5; Gaps 4;

Oy 1 tcagaaatatttgaagcctgtgtggtgaataaaagcatacaatacaatacaataca 60
Db 1391 tcagaaatatttgaagcctgtgtggtgaataaaagcatacaatacaatacaataca 1449

Oy 61 tgcataatcagccttagcaaatgagcaaaatagtaactctgttctgttctctgtct 120
Db 1450 atctaagtcagccttagcaaatgagcaaaatagtaactctgttctgttctctgtct 1509

Oy 121 acttctcagctctcaaggtcttatggccctgtgttcaactctgtattttggcctgaaac 180
Db 1510 cttttctnag-tctcaaaatctatgcccctgtgttcaactctgtattttggcctgaaac 1568

Oy 181 catagtgtgtcgtcagatgataagcagtgaggaagcctgattgattgagagaga 240
Db 1569 catgtgtgtcgtcagatgataagcagtgaggaagcctgattgattgagagaga 1628

Oy 241 gttttctggaagagcattttccacacgtgctgaaagagcctaacagagatttgtaggtg 300
Db 1629 gttttctggaagagcattttccacacgtgctgaaagagcctaacagagatttgtaggtg 1688

Oy 301 tgcattgtcctgtttcagcatctgtcttggggatggggagagatgggaaacagagacttac 360
Db 1689 tgcagtgctgtttcagcatctgtcttggggatggggagagatgggaaac--agactacg 1746

Oy 361 agagctcctcgggagagacttggcccatccacatggctggccagtgctcagcttctcttt 420
Db 1747 agagctcctcgggagagacttggcccatccacatggctggccagtgctcagcttctcttt 1806
```


OY 387 atccacatggctgccagctgcagcttctcttcttcttgcctgggatctctctctctagttt 446
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
240 atccacatggctgccagctgcagcttctcttcttgcctgggatctctctctagttt 299
OY 447 cgtttctctctctgttaggaattgtttcagcaatgaaagaaatggaaggagatccggc 506
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
300 cgtttctctctctgttaggaattgtttcagcaatgaaagaaatggaaggagatccggc 359
OY 507 gtttctctctcatgacgctgcggaattttgggatgggagagagagcatgagaccgtg 566
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
360 gtttctctctcatgacgctgcggaattttgggatgggagagagagcatgagaccgtg 419
OY 567 ttcaagaggagcccgctgccttctgagagagttgagaaaaaaccaagggtgggtgacct 626
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
420 ttcaagaggagcccgctgccttctgagagagttgagaaaaaaccaagggtgggtgacct 479
OY 627 actccatatcactgaccttactgactactatcttctctactgacattcttgaaacatt 686
Db 48^ actccatatcactgaccttactgactactatcttctctactgacattcttgaaacatt 539
OY 687 tcagggtggccatatcttcattatgagtc 717
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
540 tcagggtggccatatcttcattatgagtc 570

Search completed: April 19, 2002, 11:06:25
Job time: 13485 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 19, 2002, 09:32:08 ; Search time 4168.4 Seconds
(without alignments)
1278.330 Million cell updates/sec

Title: US-09-763-292-3

Perfect score: 323

Sequence: 1 cccctgaattgtacaacaa.....aaattcatagatcatctttt 323

Scoring table:

IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:**

1: gb_ba:**

2: gb_htg:**

3: gb_in:**

4: gb_om:**

5: gb_ov:**

6: gb_pat:**

7: gb_ph:**

8: gb_pl:**

9: gb_pr:**

10: gb_ro:**

11: gb_sts:**

12: gb_sy:**

13: gb_un:**

14: gb_vl:**

15: em_ba:**

16: em_fun:**

17: em_hum:**

18: em_in:**

19: em_om:**

20: em_or:**

21: em_ov:**

22: em_pat:**

23: em_ph:**

24: em_pi:**

25: em_ro:**

26: em_sts:**

27: em_sy:**

28: em_un:**

29: em_vl:**

30: em_hgo_hum:**

31: em_hgo_inv:**

32: em_hgo_rod:**

33: em_htg_hum:**

34: em_htg_inv:**

35: em_htg_rod:**

36: em_htg_other:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

8

Result No.	Score	Query Match	Length	DB	ID	Description
1	323	100.0	323	9	HUM2C9X05	L16881 Homo sapien
2	310.4	96.1	143087	2	AL359672	AL359672 Homo sapi
3	275.2	85.2	173154	2	AL133513	AL133513 Homo sapi
4	275.2	85.2	205791	2	AL583836	AL583836 Homo sapi
5	263	81.4	348	9	HUM2C18X06	L16874 Homo sapien
6	263	81.4	156492	2	AL157835	AL157835 Homo sapi
7	222.4	68.9	143087	2	AL359672	AL359672 Homo sapi
8	208.8	64.6	201340	2	AC018872	AC018872 Homo sapi
9	208.8	64.6	210821	2	AC013318	AC013318 Homo sapi
10	194	60.1	595	4	RABP4501	M74204 Rabbit cyto
11	192	59.4	1854	6	AR071576	AR071576 Sequence
12	192	59.4	1854	9	HUM2C9X05	M61855 Human cytoc
13	190.4	58.9	1473	22	E10631	E10631 Human cdna
14	190.4	58.9	1473	22	E10853	E10853 Human cytoc
15	190.4	58.9	1814	9	S46963	S46963 putative cy
16	190.4	58.9	1826	9	HUMCYPH2	D00173 Homo sapien
17	190.4	58.9	1843	6	E02279	E02279 Human liver
18	190.4	58.9	1845	9	HUM2C9X05	M61857 Human cytoc
19	190.4	58.9	1852	6	AR071579	AR071579 Sequence
20	188.8	58.5	1441	9	HUMCYPM8	M21940 Human cytoc
21	188.8	58.5	1576	9	HUMCYPM8	M21939 Human liver
22	188.8	58.5	1577	9	HUMCYPM8	M15331 Human cytoc
23	180.6	55.9	1892	6	AR071581	AR071581 Sequence
24	179.8	55.7	147838	2	AC083888	AC083888 Rattus no
25	179.2	55.5	1276	9	HUMCYPC17	M61858 Human cytoc
26	177.6	55.0	1346	9	HSCP450	X65962 H.sapiens m
27	177.6	55.0	1746	6	AR071575	AR071575 Sequence
28	177.6	55.0	1746	9	HUMCYPC219	M61854 Human cytoc
29	176.6	54.7	330	22	E11555	E11555 DNA probe t
30	176	54.5	1444	6	E14930	E14930 Artificial
31	176	54.5	1473	22	E10866	E10866 cdna encodi
32	176	54.5	1669	6	E14931	E14931 Human mrna
33	176	54.5	2395	9	HUMSV450A	L07093 Human cytoc
34	176	54.5	168457	2	AC078913	AC078913 Mus muscu
35	168	52.0	1473	22	E10865	E10865 cdna encodi
36	168	52.0	1995	9	HUMCYPC218	M61856 Human cytoc
37	168	52.0	2009	6	AR071577	AR071577 Sequence
38	168	52.0	2258	6	AR071580	AR071580 Sequence
39	168	52.0	2258	9	HUM2C18	M61853 Human cytoc
40	166	51.4	218	10	RATCY45M6	M18361 Rat cytochr
41	164.8	51.0	1307	4	SSSSCC103	Z93098 S.scrofa cy
42	164.8	51.0	1758	4	SSC15D09	Z93100 S.scrofa cy
43	161.2	49.9	294	10	RATP450PB1	M24237 Rat cytochr
44	158.6	49.1	1751	10	MUSCP450	D17674 Mouse mrna
45	156.8	48.5	1856	10	RATCYPM1	J02657 Rat cytochr

ALIGNMENTS

RESULT 1
HUM2C9X05
LOCUS Homo sapiens cytochrome P450C9 (CYP2C9) gene, exon 7.
DEFINITION L16881
ACCESSION L16881
VERSION 5 of 7
KEYWORDS CYP2C9; cytochrome P450; mephenytoin 4-hydroxylase.
SEGMENT 5 of 7
SOURCE human
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (sites)
AUTHORS Goldstein,J.A., Raucy,J.L., Blaisdell,J.A., Faletto,M.B. and Romkes,M.
TITLE Cloning and expression of complementary DNAs for multiple members of the human cytochrome P4501C subfamily
JOURNAL Biochemistry 30, 3247-3255 (1991)
MEDLINE 91182740
REFERENCE 2 (bases 1 to 323)

AUTHORS de Morais, S.M., Schweikl, H., Blaisdell, J. and Goldstein, J.A.
 TITLE Gene structure and upstream regulatory regions of human CYP2C9 and
 CYP2C18
 JOURNAL Biochem. Biophys. Res. Commun. 194 (1), 194-201 (1993)
 MEDLINE 93326116
 FEATURES

source Location/Qualifiers
 1..323
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /tissue_type="liver"
 /tissue_lib="EMBL3 library from J.A. Goldstein"
 70..257
 /gene="CYP2C9"
 /note="coding sequence identical to Lue 359 variant (clone 25)"
 /citation=[1]
 /number=7
 /evidence-experimental 83 a 88 c 59 g 93 t

BASE COUNT 83 a 88 c 59 g 93 t
 ORIGIN
 Query Match 100.0%; Score 323; DB 9; Length 323;
 Best Local Similarity 100.0%; Pred. No. 1.5e-84;
 Matches 323; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 cccctgaattgctacacaaatgtgccattttctctctttccatcagtttttacttg 60
 Db 1 CCCCTGAATTCCTACACAAATGTGCCATTTCTCTTTTCCATCAGTTTTACTTTG 60
 Oy 61 tcttatcagctaaagtcagaaagattgaacgtgtgattggcagaaacccgagccct 120
 Db 61 TCTTATCAGCTAAAGTCAGGAAGAGATTGAAGCTGTGATTGGCAGAAACCGAGCCCT 120
 Oy 121 gcatgcaagacagagccacatgccctacacagatgctggtgacagagtcagagat 180
 Db 121 GCATGCAAGACAGGAGCCACATGCCCTACACAGATGCTGTGGTGCAGAGTCCAGAT 180
 Oy 181 accttgaccttctccaccagcctgcccacatgcagtgacctgtgacattaaattcagaa 240
 Db 181 ACCTTGACCTTCTCCACCAGCGCTGCCCATGCAGTGCAGCTGTGACATTAAATTCAGAA 240
 Oy 241 actatctcattcccaagtgattgtttctctcactacagcgaactcagtttccgaagt 300
 Db 241 ACTATCTCATTCCAAGTAAGTTGTTTCTCTCCTACACTGCAACTCCATGTTTCGAAGT 300
 Oy 301 cccaaattcatagatcattttt 323
 Db 301 CCCAAATTCATAGTATCATTTT 323

RESULT 2
 LOCUS AL359672 143087 bp DNA HTG 27-JUN-2001
 DEFINITION Homo sapiens chromosome 10 clone RP11-208C17, *** SEQUENCING IN
 PROGRESS ***, 2 unordered pieces.
 ACCESSION AL359672
 VERSION AL359672.17 GI:14575223
 KEYWORDS HTG: HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
 SOURCE human.
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 143087)

Johnson, C.
 Direct Submission
 Submitted (26-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire,
 CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
 requests: clonerequest@sanger.ac.uk
 On Jun 28, 2001 this sequence version replaced gi:14529836.
 ----- Genome Center
 Center: Sanger Centre
 Center code: SC

COMMENT

Web site: <http://www.sanger.ac.uk>
 Contact: humquery@sanger.ac.uk
 ----- Project Information
 Center project name: BA208C17
 ----- Summary Statistics
 Assembly program: XGAP4; version 4.5
 Sequencing vector: pLasmid; L08752; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Consensus quality: 142666 bases at least Q40
 Consensus quality: 142843 bases at least Q30
 Consensus quality: 142920 bases at least Q20
 Insert size: 142987; sum-of-contigs
 Insert size: 156404; agarose-fp
 Quality coverage: 8.55x in Q20 bases; sum-of-contigs Quality
 coverage: 7.82x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 2 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1 47005: contig of 47005 bp in length
 * 47006 47105: gap of 100 bp
 * 47106 143087: contig of 95982 bp in length.

FEATURES

source

1..143087
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="10"
 /clone="RP11-208C17"
 /clone_lib="RPC1-11.1"
 /note="assembly_fragment:00494
 clone_end:SP6
 vector_side:left"

misc_feature

1..47005

/note="assembly_fragment:00967"
 47106..143087

misc_feature

47106..143087
 /note="assembly_fragment:00967"
 26155 c 27788 g 47797 t 100 others

BASE COUNT

ORIGIN

Query Match 96.1%; Score 310.4; DB 2; Length 143087;
 Best Local Similarity 99.4%; Pred. No. 1.4e-80;
 Matches 322; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

Oy 1 cccctgaattgctacacaaatgtgccattttctctctttccatcagtttttacttg 60
 Db 99506 CCCCTGAATTCCTACACAAATGTGCCATTTCTCTTTTCCATCAGTTTTACTTTG 60
 Oy 61 tcttatcagctaaagtcagaaagattgaacgtgtgattggcagaaacccgagccct 120
 Db 99566 TCTTATCAGCTAAAGTCCAGGAAGAGATTGAAGCTGTGATTGGCAGAAACCGAGCCCT 99625
 Oy 121 gcatgcaagacagagccacatgccctacacagatgctggtgacagagtcagagat 180
 Db 99626 GCATGCAAGACAGGAGCCACATGCCCTACACAGATGCTGTGGTGCAGAGTCCAGAT 99685
 Oy 181 accttgaccttctccaccagcctgcccacatgcagtgacctgtgacattaaattcagaa 240
 Db 99686 ACATGTAGCTTCTCCACCAGCGCTGCCCATGCAGTGCAGTGCAGTGCAGTGCAGT 99745
 Oy 241 actatctcattcccaagtgattgtttctctcactacagcgaactcagtttccgaagt 300
 Db 99746 ACTATCTCATTCCTCAAGTAAGTTGTTTCTCTCCTACACTGCAACTCCATGTTTCGAAGT 99805
 Oy 301 cccaaattcatagatcattttt 323
 Db 99806 CCCAAATTCATAGTATCATTTT 99829

RESULT	3
LOCUS	AL133513 173154 bp DNA HTG 04-JUL-2001
DEFINITION	Homo sapiens chromosome 10 clone RP11-400G3, *** SEQUENCING IN PROGRESS ***, in ordered pieces.
ACCESSION	AL133513
VERSION	AL133513.11 GI:14575067
KEYWORDS	HTG; HTGS_PHASE2; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	Brown,J.
TITLE	Direct Submission
JOURNAL	Submitted (03-JUL-2001) Sanger Centre, Hinxtan, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk requests: clonerequest@sanger.ac.uk On Jun 28, 2001 this sequence version replaced gi:14456168.
COMMENT	----- Genome Center Center: Sanger Centre Center code: SC Web site: http://www.sanger.ac.uk Contact: humquery@sanger.ac.uk ----- Project Information Center project name: BA400G3 ----- Summary Statistics Sequencing program: XGAP4; version 4.5 Chemistry: Dye-terminator Big Dye; 100% of reads Consensus quality: 172989 bases at least Q40 Consensus quality: 173136 bases at least Q20 Consensus quality: 173145 bases at least Q20 Insert size: 173154; sum-of-contigs Insert size: 116005; 33.1% error; agarose-fp Quality coverage: 6.48x in Q20 bases; sum-of-contigs Quality coverage: 9.80x in Q20 bases; agarose-fp ----- * NOTE: This is a 'working draft' sequence. * This sequence will be replaced * by the finished sequence as soon as it is available and * the accession number will be preserved.
FEATURES	Location/Qualifiers
source	1..173154 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="10" /clone="RP11-400G3" /clone_lib="RPC1-11.2" 1..173154 /note="assembly_fragment:00748 clone_end:sp6 vector_side:right"
BASE COUNT	49163 a 32272 c 33707 g 58012 t
ORIGIN	
Query Match	85.2%; Score 275.2; DB 2: Length 173154;
Best Local Similarity	92.6%; Pred. No. 3.4e-70;
Matches 300; Conservative	0; Mismatches 23; Indels 1; Gaps 1
QY	1 ccctgaattgctacaaagtgcgaagatgtgccatttttctcttttcacagttttacttgatg 60
Db	51799 CCCTGAATTGCTAGAACAAATGTTCATTTTCTCTCTTTTCCCATCACTTGCTACTTG 51858
QY	61 ttcttatcagctaaaagtcaggagagatgtaacgltgtgatlgggcagaaccgagccct 120
Db	51859 TCTTGTCACTAAAGTCCAGGAAGATTGAACGTGTCCTGGCANAACC GGAGCCCC 51918
QY	121 gcatgcaagacaggagccacatgccctcacagatgttgtgtggtcgacaggtccagagat 180
Db	51919 GCATGCAGCACAGGGGCCACATGCCCTACACAGATGCTGTGGTGACAGGTCCAGAGAT 51978

/note="assembly_fragment:01703
fragment_chain:1
clone_end:T7
vector_side:right"

BASE COUNT 56694 a 38262 c 40890 g 69845 t 100 others
ORIGIN

Query Match 85.2%; Score 275.2; DB 2; Length 205791;
Best Local Similarity 92.6%; Pred. No. 3.4e-70;
Matches 300; Conservative 0; Mismatches 23; Indels 1; Gaps 1;
Qy 1 cccctgaattgctacaacaaatgctccatttttctctttccatcagtttttactgtg 50
Db 164775 CCCCTGAATTGCTAGAACAAATGTCATTTCCTCTTTTCCATCAGTCTTACTTGG 164834
Qy 61 tcttatcagctaaagtccagaaagagattgaacgtgtgattgagcagaaacccgagccct 120
Db 164835 TCTTGTGACGTAAAGTCAGGAAGAGATTGAACGTGCTGTTGGCAGAAACCGAGCCCT 164894
Qy 121 gcatgcaagacagagccacatgccctacacagatgctgtggtgcacagagtcagagat 180
Db 164895 GCATGCAGGACAGGGCCACATGCCCTACACAGATGCTGTGGTGCACAGGTCAGAGAT 164954
Qy 181 acctgacctctcccacacagcctgcccacatgcagtgacctgtgacattaaattcagaa 240
Db 164955 ACATCGACCTCATCCCAACAGCCTGCCCATGCAGTGACCTGTGACGTTAAATTCAGAA 165014
Qy 241 actatctcattcccaagtaagttgtttctctacactgcaactcactcatctgttcgaagt 300
Db 165015 ACTACTCATTCCTCCAGGTAAGTTGTTCTCTCTACACTGCAACTCCTCATGTTCTTTAT 165074
Qy 301 cc-caaattcattagatcattttt 323
Db 165075 CCTCAAAATTCACAGATGATGTTCT 165098

RESULT 5
LOCUS HUM2C18X06
DEFINITION Homo sapiens cytochrome P4502C18 (CYP2C18) gene, exon 7.
ACCESSION L16874
VERSION L16874.1 GI:291604
KEYWORDS CYP2C18; cytochrome P450; mephenytoin 4-hydroxylase.
SEGMENT 6 of 8
SOURCE Homo sapiens (library: EMBL3 library from J.A. Goldstein) liver
DNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (sites)
AUTHORS Goldstein,J.A., Raucy,J.L., Blaisdell,J.A., Faletto,M.B. and Romkes,M.
TITLE Cloning and expression of complementary DNAs for multiple members of the human cytochrome P4501C subfamily
JOURNAL Biochemistry 30, 3247-3255 (1991)
MEDLINE 91182740
REFERENCE 2 (bases 1 to 348)
AUTHORS de Morais,S.M., Schweikl,H., Blaisdell,J. and Goldstein,J.A.
TITLE Gene structure and upstream regulatory regions of human CYP2C9 and CYP2C18
JOURNAL Biochem. Biophys. Res. Commun. 194 (1), 194-201 (1993)
MEDLINE 93326116
REFERENCE 3 (sites)
AUTHORS Romkes,M., Faletto,M.B., Blaisdell,J.A., Raucy,J.L. and Goldstein,J.A.
TITLE Correction: Cloning and expression of complementary cDNAs for multiple members of the human cytochrome P4501C subfamily
JOURNAL Biochemistry 32, 1390-1390 (1993)
MEDLINE 93192243
FEATURES Location/Qualifiers
source 1..348,
/organism="Homo sapiens"

/db_xref="taxon:9606"
/tissue_type="liver"
/tissue_lib="EMBL3 library from J.A. Goldstein"
order(L16873,1:206..>325,<1..100)
/gene="CYP2C18"
/number=6
101..288
/gene="CYP2C18"
/citation={1}
/citation={3}
/number=7
/evidence-experimental
BASE COUNT 89 a 97 c 62 g 100 t
ORIGIN

Query Match 81.4%; Score 263; DB 9; Length 348;
Best Local Similarity 91.7%; Pred. No. 7.2e-67;
Matches 289; Conservative 0; Mismatches 25; Indels 1; Gaps 1;
Qy 3 cctgaattgctacaacaaatgctccatttttctctttccatcagtttttactgtg 62
Db 34 CCTGAATTGCTACACAAATGTGCCATTTTCCTCTTTTCCATCATTTCTTACTTGTGC 93
Qy 53 ttatcagctaaagtccagaaagagattgaacgtgtgattgagcagaaacccgagccctgc 122
Db 94 TTATCAGCTAAAGTCAGGAAGAGATTGAATGTTGTCAGTGTGTCGAGAAACCGGAGCCCTGT 153
Qy 123 atgcaagacagagcagcagatgccctacacagatgctgtggtgcacagagtcagagatac 182
Db 154 ATGCAGGACAGGAGTACATGCCCTACACAGATGCTGTGGTGCACGAGATCCAGAGATAC 213
Qy 183 ctgacctctcccacacagcctgcccacatgcagtgacctgtgacattaaattcagaaac 242
Db 214 ATGACCTCTCTCCCAACCACTGCCCATGCAGTGACCTGTGATGTTAAATTCAAAAAC 273
Qy 243 tatctcattcccaagtaagttgtttctctacactgcaactcactcatctgttcgaagt-c 301
Db 274 TACCTATCCCCAAGTAGCTTGTTCCTCTACACTACATCTCCATCTCCTCAAGTCC 333
Qy 302 ccaaattcattagat 316
Db 334 CCAAATTCATAGTAT 348

RESULT 6
LOCUS AL157835 156492 bp DNA
DEFINITION Homo sapiens chromosome 10 clone RP11-361K9, *** SEQUENCING IN PROGRESS ***, in ordered pieces.
ACCESSION AL157835
VERSION AL157835.9 GI:14626943
KEYWORDS HTG: HTGS_PHASE2; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 156492)
AUTHORS Blakey,S.
TITLE Direct Submission
JOURNAL Submitted (05-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK
E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Jul 8, 2001 this sequence version replaced gi:14586042.
COMMENT
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA361K9
----- Summary Statistics
Assembly program: XGAP4; version 4.5

```

Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 156104 bases at least Q40
Consensus quality: 156214 bases at least Q30
Consensus quality: 156256 bases at least Q20
Insert size: 156492; sum-of-contigs
Insert size: 150975; 13.7% error; agarose-1p
Quality coverage: 8.34x in Q20 bases; sum-of-contigs Quality
coverage: 9.69x in Q20 bases; agarose-1p
-----
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES             Location/Qualifiers
     source            1..156492
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /chromosome="10"
                        /clone="RP11-361K9"
                        /clone_lib="RP11-11.2"
                        /note="assembly_fragment:03685
                        clone_end:SP6
                        vector_side:left"
BASE COUNT           43071 a 30007 c 31050 g 52364 t
ORIGIN
Query Match          81.4%; Score 263; DB 2; Length 156492;
Best Local Similarity 91.7%; Pred. No. 1.3e-66;
Matches 289; Conservative 0; Mismatches 25; Indels 1; Gaps 1;

Oy 3 cctgaattgtacacaaatgtgccattttctcttccatcagttttacttgctc 62
|||||
Db 140291 CCTGAATTGCTACAAATGTGCCATTTCCTCTTTCATCAITTCITACITGTGC 140350

Oy 63 ttatcagcttaagtccaggaagatgtgaactgtgtggtggcagaaaccgagccctgc 122
|||||
Db 140351 TTATCAGCTTAAAGTCCAGGAAGAGATTGAATGTGTAGTTGGCAGAAACCGGAGCCCTGT 140410

Oy 123 atgcaagacagagccacatgcctcacagagatgctgtggtgacagaggtccagagatac 182
|||||
Db 140411 ATGCAGGACAGAGTACATGCCCTTACACAGATGCTGTGGTGCAGGATCCAGAGATAC 140470

Oy 183 ctgtgacctctcccaccagcctgccccatgcagtgacctgtgacattaaattcagaaac 242
|||||
Db 14047; ATTGACCTCTCCCCACCACCACTGCCCATGCAGTGACCTGTGATGTTAAATTCAAAAC 140530

Oy 243 tatctcattcccaagtgtaagttgttctctacactgcaactccatgttttcgaagt-c 301
|||||
Db 140531 TACCTCATCCCAAGTAAAGTGTGTTCTCTTACACTACATCTCCATGCTCTTCAAGTCC 140590

Oy 302 ccaaatccatagtat 316
|||||
Db 140591 CCAAAATTCATAGTAT 140605

RESULT 7
AL359672/c
LOCUS
DEFINITION Homo sapiens chromosome 10 clone RP11-208C17, *** SEQUENCING IN
PROGRESS ***, 2 unordered pieces.
ACCESSION AL359672
VERSION AL359672.17 GI:14575223
KEYWORDS HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 143087)
AUTHORS Johnson,C.
TITLE Direct Submission

```

JOURNAL

COMMENT

Submitted (26-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
 On Jun 28, 2001 this sequence version replaced gi:14529836.
 ----- Genome Center
 Center: Sanger Centre
 Center code: SC
 Web site: http://www.sanger.ac.uk
 Contact: humquery@sanger.ac.uk
 ----- Project Information
 Center project name: BA208C17
 ----- Summary Statistics
 Sequencing program: XGAP4; version 4.5
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Consensus quality: 142666 bases at least Q40
 Consensus quality: 142843 bases at least Q30
 Consensus quality: 142920 bases at least Q20
 Insert size: 142987; sum-of-contigs
 Insert size: 156404; agarose-1p
 Quality coverage: 8.55x in Q20 bases; sum-of-contigs Quality coverage: 7.82x in Q20 bases; agarose-1p

* NOTE: This is a 'working draft' sequence. It currently consists of 2 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

* 1 47005: contig of 47005 bp in length
 * 47006 47105: gap of 100 bp
 * 47106 143087: contig of 95982 bp in length.

FEATURES

source

1..143087
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="10"
 /clone="RP11-208C17"
 /clone_lib="RP11-11.1"
 /note="assembly_fragment:00494
 vector_side:left"
 47106..143087
 /note="assembly_fragment:00967"

misc_feature

1..47005
 /note="assembly_fragment:00494
 vector_side:left"
 47106..143087
 /note="assembly_fragment:00967"

BASE COUNT

41247 a 26155 c 27788 g 47797 t 100 others

ORIGIN

Query Match 68.9%; Score 222.4; DB 2; Length 143087;
 Best Local Similarity 83.2%; Pred. No. 1.2e-54;
 Matches 253; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

Oy 2 cctgaattgtacacaaatgtgccattttctcttccatcagttttacttgctg 61
 |||
 Db 18256 CCATGAATTGCTATGACAAATGTTCCATATATCTTCTTCCATCAGTCTTCTTGTGT 18197

Oy 62 ctatcagcttaagtccaggaagatgtgaactgtgtgacgtgattggcagaaaccgagccctg 121
 |||
 Db 18196 CTGTGACGCTAAAGTCCAGGAAGAGATTGATCATGTAATTGGCAGACACAGGAGCCCTG 18137

Oy 122 catgcaagacagagccacatgcctcacagatgctgtggtgacagaggtccagagata 181
 |||
 Db 18136 CATGCAGGATAGGAGCCACATGCCTTACTACTGATGCTGTAGTGCACGAGATCCAGAGATA 18077

Oy 182 ccttgacctctcccaccagcctgccccatgcagtgacctgtgacattaaattcagaaa 241
 |||
 Db 18076 CAGTGACCTTGTGCCCCCGGTGTCGCCCATGCAGTGACCACTGATACTAAGTTCAGAAA 18017

Oy 242 ctatctcattcccaaggttaagttgttctctacactgcaactccatggttttcgaagtc 301
 |||

Db 18016 CTACCTATCCCAAGGTAAGCTGTGTTCTCTTACTATATTTCTGACTTCTGAAT 17957

RESULT	8
AC018872	
LOCUS	AC018872 201340 bp DNA HTG
DEFINITION	Homo sapiens chromosome 2 clone RP11-140M22, WORKING DRAFT
ACCESSION	AC018872
VERSION	AC018872.6 GI:12963039
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE	human.
ORGANISM	Homo sapiens

SOURCE	ORGANISM	REFERENCE
Human.	Homo sapiens	
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
REFERENCE		1 (bases 1 to 201340)
AUTHORS	Waterston, R.H.	
TITLE	The sequence of Homo sapiens clone	
JOURNAL	Unpublished	
REFERENCE		2 (bases 1 to 201340)
AUTHORS	Waterston, R.H.	
TITLE	Direct Submission	
JOURNAL	Submitted (21-DEC-1999) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis MO 63108, USA	
COMMENT	On Feb 18, 2001 this sequence version replaced gi:89544228.	

```

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH0140M22
----- Summary Statistics -----
Sequencing vector: M13; 71k
Chemistry: Dye-primer; 29k
Chemistry: Dye-terminator; Big Dye; 29k of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 187240 bases at least Q40
Consensus quality: 190955 bases at least Q30
Consensus quality: 193344 bases at least Q20
Insert size: 178000; agarose-fp
Insert size: 198640; sum-of-contigs
Quality coverage: 4.00 in Q20 bases; agarose-fp
Quality coverage: 4.13 in Q20 bases; sum-of-contigs

```

*	1	1402:	contig of 1402 bp in length
*	1403	1502:	gap of unknown length
*	1503	2841:	contig of 1339 bp in length
*	2842	2941:	gap of unknown length
*	2942	4465:	contig of 1524 bp in length
*	4466	4565:	gap of unknown length
*	4566	7341:	contig of 2776 bp in length
*	7342	7441:	gap of unknown length
*	7442	9248:	contig of 1807 bp in length
*	9249	9348:	gap of unknown length
*	9349	11809:	contig of 2461 bp in length
*	11810	11909:	gap of unknown length
*	11910	15022:	contig of 3113 bp in length
*	15023	15123:	gap of unknown length

* 15123	18880: contig of 3758 bp in length
* 18881	18980: gap of unknown length
* 18981	21664: contig of 2684 bp in length
* 21665	21764: gap of unknown length
* 21765	25032: contig of 3268 bp in length
* 25033	25132: gap of unknown length
* 25133	30071: contig of 4939 bp in length
* 30072	30171: gap of unknown length
* 30172	35416: contig of 5245 bp in length
* 35417	35516: gap of unknown length
* 35517	40752: contig of 5236 bp in length
* 40753	40852: gap of unknown length
* 40853	45905: contig of 5053 bp in length
* 45906	46005: gap of unknown length
* 46006	50784: contig of 4779 bp in length
* 50785	50884: gap of unknown length
* 50885	56904: contig of 6020 bp in length
* 56905	57004: gap of unknown length
* 57005	64526: contig of 7522 bp in length
* 64527	64626: gap of unknown length
* 64627	72095: contig of 7469 bp in length
* 72096	72195: gap of unknown length
* 72196	79372: contig of 7177 bp in length
* 79373	79472: gap of unknown length
* 79473	87932: contig of 8460 bp in length
* 87933	88032: gap of unknown length
* 88033	97114: contig of 9082 bp in length
* 97115	97214: gap of unknown length
* 97215	10392: contig of 12178 bp in length
* 10393	109492: gap of unknown length
* 109493	119347: contig of 9855 bp in length
* 119348	119447: gap of unknown length
* 119448	133975: contig of 14528 bp in length
* 133976	134075: gap of unknown length
* 134076	150216: contig of 16141 bp in length
* 150217	150316: gap of unknown length
* 150317	164328: contig of 14012 bp in length
* 164329	164428: gap of unknown length
* 164429	178474: contig of 14046 bp in length
* 178475	178574: gap of unknown length
* 178575	201340: contig of 22766 bp in length

FEATURES

Location/Qualifiers

1. .201340

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="2"

/clone="RP11-140M22"

misc_feature

1. .1402

/note="assembly_name:Contig11"

misc_feature

1503. .2841

/note="assembly_name:Contig12"

2942. .4465

/note="assembly_name:Contig13"

4566. .7341

/note="assembly_name:Contig15"

7442. .9248

/note="assembly_name:Contig16"

9349. .11809

/note="assembly_name:Contig17"

11910. .15022

/note="assembly_name:Contig18"

15123. .18880

/note="assembly_name:Contig19"

18981. .21664

/note="assembly_name:Contig21

clone_end:17

vector_side:left

21765. .25032

/note="assembly_name:Contig20"

25133. .30071

/note="assembly_name:Contig22"

30172. .35416

/note="assembly_name:Contig23"

```

misc_feature      35517..40752
/note="assembly_name:Contig24"
misc_feature      40853..45905
/note="assembly_name:Contig25"
misc_feature      46006..50784
/note="assembly_name:Contig26"
misc_feature      50885..56904
/note="assembly_name:Contig27"
misc_feature      57005..64526
/note="assembly_name:Contig28"
misc_feature      64627..72095
/note="assembly_name:Contig29"
misc_feature      72196..79372
/note="assembly_name:Contig30"
misc_feature      79473..87932
/note="assembly_name:Contig31"
misc_feature      88033..97114
/note="assembly_name:Contig32"
misc_feature      97215..109392
/note="assembly_name:Contig33"
misc_feature      109493..119347
/note="assembly_name:Contig34"
misc_feature      119448..133975
/note="assembly_name:Contig35"
misc_feature      134076..150216
/note="assembly_name:Contig36"
misc_feature      150317..164328
/note="assembly_name:Contig37"
misc_feature      164429..178474
/note="assembly_name:Contig38"
misc_feature      178575..201340
/note="assembly_name:Contig39"

BASE COUNT      61711 a 36248 c 35995 g 64637 t 2749 others
ORIGIN

Query Match      64.6%; Score 208.8; DB 2: Length 201340;
Best Local Similarity 82.5%; Pred. No. 1.3e-50;
Matches 264; Conservative 0; Mismatches 52; Indels 4; Gaps 2;

QY 3 cctgaattgtacaaataatgtccattttctctttccatcagttttactgtgtc 62
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 116688 CCTGAATGCTACGACAAATGTGCCATTTTCTCCCTTTTCATCAGCTCCACTGTGTC 116747

QY 63 ttatcagctaaagtccagggaagattgacgtgtgtattgacagaaaccggacccttc 122
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 116748 TTGTCATCTAAAGTCCAGGAGAGACGCGACCATGTCAGTTGGCAGACTGGAGCCCTGC 116807

QY 123 atgcaagacaggagccacatgccctacacagatgctgtgtgacaggtccagagatac 182
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 116808 ATGCAGGACAGGAGCCACATGCCCTACACAGAAGCCATGGTGCATGAGTCCAGAGACAC 116867

QY 183 cttagaccttccccaccagcctgccccatgcagtcagtcgactgaacattacagaacac 242
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 116868 ---TGACCTACCCCAATGTCGCCCATGTCAGTACCTGACCTGACATTAATTTAGAAAC 116924

QY 243 tatctattcccaaggttaattgtttctctctacactgcaactccatttttcgaagt-c 301
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 116925 TACCTCTCCCAAGGTAAGCTTGTTCCTCTAGACTGTCGCTCTGCTCTTGAATGCC 116984

QY 302 ccaaatcagatcatcttt 321
||||| ||||| ||||| |||||
Db 116985 CCAAAATTCAGATTAATGTTT 117004

RESULT 9
AC013318/c
LOCUS AC013318 210821 bp DNA HTG 23-SEP-2000
DEFINITION Homo sapiens chromosome 2 clone RP11-358N5 map 2, WORKING DRAFT
SEQUENCE, 22 unordered pieces.
ACCESSION AC013318
VERSION AC013318.5 GI:10280844
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.

```

SOURCE ORGANISM

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 210821)

Birren, B., Linton, L., Nusbaum, C. and Lander, E.

Homo sapiens chromosome 2, clone RP11-358N5

TITLE

Unpublished

JOURNAL

2 (bases 1 to 210821)

AUTHORS

Baldwin, J., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,

Brown, A., Castle, A., Colangelo, M., Collins, S., Collymore, A.,

Cooke, P., DeArelano, K., Dewar, K., Domino, M., Donelan, L., Doyle, M.,

Ferreira, P., FitzHugh, W., Forrest, C., Funke, R., Gage, D.,

Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,

Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,

Lehoczy, J., Lieu, C., Locke, K., Macdonald, P., Marquis, N.,

McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrim, J.,

Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,

Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,

Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,

Tesfaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,

Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.

Direct Submission

Submitted (06-NOV-1999) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

On Sep 23, 2000 this sequence version replaced gi:8072541.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L3773

Center clone name: 358_N_5

----- Summary Statistics

Sequencing vector: M13; M77815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 196258 bases at least Q40

Consensus quality: 203137 bases at least Q30

Consensus quality: 206079 bases at least Q20

Insert size: 196000; agarose-fp

Quality coverage: 4.7 in Q20 bases; agarose-fp

Quality coverage: 4.4 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently

* consists of 22 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

* 1 5808: contig of 5808 bp in length

* 5809 5908: gap of 100 bp

* 5909 7108: contig of 1200 bp in length

* 7109 7208: gap of 100 bp

* 7209 8378: contig of 1170 bp in length

* 8379 8478: gap of 100 bp

* 8479 9873: contig of 1395 bp in length

* 9874 9973: gap of 100 bp

* 9974 12585: contig of 2612 bp in length

* 12586 12685: gap of 100 bp

* 12686 16506: contig of 3821 bp in length

* 16507 16606: gap of 100 bp

* 16607 20660: contig of 4054 bp in length

* 20661 20760: gap of 100 bp

* 20761 24298: contig of 3538 bp in length

```

* 24299 24398: gap of 100 bp
* 24399 29876: contig of 5478 bp in length
* 29877 29976: gap of 100 bp
* 29977 36607: contig of 6631 bp in length
* 36608 36707: gap of 100 bp
* 36708 43142: contig of 6435 bp in length
* 43143 43242: gap of 100 bp
* 43243 49210: contig of 5968 bp in length
* 49211 49310: gap of 100 bp
* 49311 55880: contig of 6570 bp in length
* 55881 55980: gap of 100 bp
* 55981 63850: contig of 7870 bp in length
* 63851 63950: gap of 100 bp
* 63951 71005: contig of 7055 bp in length
* 71006 71105: gap of 100 bp
* 71106 80701: contig of 9596 bp in length
* 80702 80801: gap of 100 bp
* 80802 126993: contig of 46192 bp in length
* 126994 127093: gap of 100 bp
* 127094 139959: contig of 12866 bp in length
* 139960 140059: gap of 100 bp
* 140060 155880: contig of 15821 bp in length
* 155881 155980: gap of 100 bp
* 155981 170719: contig of 14739 bp in length
* 170720 170819: gap of 100 bp
* 170820 200516: contig of 29697 bp in length
* 200517 200616: gap of 100 bp
* 200617 210821: contig of 10205 bp in length.
FEATURES
    source
        Location/Qualifiers
            1..210821
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /map="2"
                /clone="RP11-358N5"
                /clone_lib="RP11 Human Male BAC"
            1..5808
                /note="assembly_fragment"
                clone_end:SP6
                vector_side:left"
                5909..7108
                    /note="assembly_fragment"
                    7209..8378
                        /note="assembly_fragment"
                        8479..9873
                            /note="assembly_fragment"
                            9974..12585
                                /note="assembly_fragment"
                                12686..16506
                                    /note="assembly_fragment"
                                    16607..20660
                                        /note="assembly_fragment"
                                        20761..24298
                                            /note="assembly_fragment"
                                            24399..29876
                                                /note="assembly_fragment"
                                                29977..36607
                                                    /note="assembly_fragment"
                                                    36708..43142
                                                        /note="assembly_fragment"
                                                        43243..49210
                                                            /note="assembly_fragment"
                                                            49311..55880
                                                                /note="assembly_fragment"
                                                                55981..63850
                                                                    /note="assembly_fragment"
                                                                    63951..71005
                                                                        /note="assembly_fragment"
                                                                        71106..80701
                                                                            /note="assembly_fragment"
                                                                            80802..126993
                                                                                /note="assembly_fragment"
                                                                                127094..139959

```

```

misc_feature
    /note="assembly_fragment"
    140060..155880
misc_feature
    /note="assembly_fragment"
    155981..170719
misc_feature
    /note="assembly_fragment"
    170820..200516
misc_feature
    /note="assembly_fragment"
    200617..210821
    /note="assembly_fragment"
    clone_end:T7
    vector_side:right"
BASE COUNT 67328 a 37637 c 36968 g 66777 t 2111 others
ORIGIN

Query Match 64.6% Score 208.8; DB 2; Length 210821;
Best Local Similarity 82.5%; Pred. No. 1.3e-50;
Matches 264; Conservative 0; Mismatches 52; Indels 4; Gaps 2;

Qy 3 cctgaattgtacacaaatgcccattttctcttctccatcagtttttacttgtgc 62
Db 25158 CCTGAAGTGTACGACAAATGGCCATTTTCTCCTTTTTCATCAGTCTCACTTGTGTC 25099

Qy 63 ttatcagctaaagtccagggaagatgaaagtgtgattgacagaaacggagccctgc 122
Db 25098 TTGTCATCTAAAGTCCAGGAGAGACGACGACATGCGTGGCAGACACTGGAGCCCTGC 25039

Qy 123 atcgaacagaggagcccatgcccacacagatgctgacacagtgctgacacaggtccagagatac 182
Db 25038 ATGCAGGACAGGAGCCACATGCCCTACACAGAGCCATGTCATGATGAGGTCAGAGACAC 24979

Qy 183 ctgacacctctcccccagccgctgcccagtcgagtgacacctgtgacattaaattcagaaac 242
Db 24978 ---TGACCTCACCCCAATGTGCCCATGTGCTGACCTGACCTCTGACATTAATTTAGAAAC 24922

Qy 243 tatctcattcccaaggttaagttgtttctctcacactgcaactcctatgttttcgaagt-c 301
Db 24921 TACCTCTCTCCCAAGGTAAAGCTTGTTCCTCTAGACTGTGCGTCTATGCTCTTGATGTC 24862

Qy 302 ccaaaattcatagtcatatt 321
Db 24861 CCAAAATTCACAGTATTGTTT 24842

RESULT 10
RABP4501
LOCUS
    RABP4501 595 bp DNA MAM 27-APR-1993
DEFINITION
    Rabbit cytochrome P45011C5 (Cyp2C5) gene, exon 7.
ACCESSION
    M74204
VERSION
    M74204.1 GI:165568
KEYWORDS
    cytochrome P450 IIC5; drug metabolism; microsomal membrane protein;
    monooxygenase.
SEGMENT
    1 of 3
SOURCE
    Oryctolagus cuniculus (strain New Zealand White) adult liver DNA.
ORGANISM
    Oryctolagus cuniculus
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
REFERENCE
    1 (bases 1 to 595)
    Zhao,J., Chan,G., Govind,S., Bell,P. and Kemper,B.W.
    Structure of 5' regions and expression of phenobarbital-inducible
    rabbit cytochrome P45011C genes
    DNA Cell Biol. 9, 37-48 (1990)
JOURNAL
    90197893
MEDLINE
    Location/Qualifiers
FEATURES
    source
        1..595
            /organism="Oryctolagus cuniculus"
            /strain="New Zealand White"
            /db_xref="taxon:9986"
            /cell_type="hepatocyte"
            /tissue_type="liver"
            /dev_stage="adult"
            /germline
            162..349

```

exon


```
/gene="CYP2C5"
/number=7
/evidence=experimental
BASE COUNT 160 a 130 c 110 g 195 t
ORIGIN

Query Match 60.1%; Score 194; DB 4; Length 595;
Best Local Similarity 81.0%; Pred. No. 1.7e-46;
Matches 238; Conservative 0; Mismatches 55; Indels 1; Gaps 1;

Oy 2 cccatgaattgctacacaaatgcccattttctctctt-tccatcagtttttactttgtg 60
||||| ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 93 CCTAAATGTATGACAGCTGTGCCAATTTCTCAITTCCTCCGTTATTCTTACTTAIG 152

Oy 61 tcttatcagctaaagtccagaagaattgaacgtgtgattggcagaaaccggagccct 120
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 153 TCTTATCAGCTAGGTCCTCAGAGAGATTGAGCGTGTGATTGGCAGACACGGAGCCCC 212

Oy 121 gcatgaagacaggagccacatgccctacacagatgctgtgtgacagaggtccagat 180
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 213 GCATGCAGACAGGAGCGCATGCCCTACACGGATGCTGTGATACATGAGATCCAGAGAT 272

Oy 181 acctgaacttctcccacagcctgcccctgacgtgacgtgacattaaattcagaa 240
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 273 TCATCGACCTCTCTCCCACTAACCTGCCCATGCCATGACGTGACCCGAGATGTAGATTGAGAA 332

Oy 241 actatctcattcccaagtaagtgtttctcctacactgaactccatgtttt 294
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 333 ACTACTTTATCCCTAAGGTATCTGTTCTTTTACATGACCTCAGTGTCT 386

RESULT 11
AR071576 1854 bp DNA PAT 18-FEB-2000
LOCUS AR071576 Sequence 4 from patent US 5912120.
DEFINITION AR071576
ACCESSION AR071576
VERSION AR071576.1 GI:7222464
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1854)
AUTHORS Goldstein,J.A. and De Morais,S.M.F.
TITLE Cloning, expression and diagnosis of human cytochrome P450 2C19:
the principal determinant of s-mephenytoin metabolism
JOURNAL Patent: US 5912120-A 4 15-JUN-1999;
FEATURES
source Location/Qualifiers
1..1854
/organism="unknown"
BASE COUNT 513 a 424 c 381 g 536 t
ORIGIN

Query Match 59.4%; Score 192; DB 6; Length 1854;
Best Local Similarity 100.0%; Pred. No. 7.3e-46;
Matches 192; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 67 cagctaaagtccaggaagattgaacgtgtgattggcagaaccggagccctgcagtc 126
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 971 CAGCTAAAGTCCAGGAAGAGATTGAACGTGTGATTGGCAGAAACCGGAGCCCTGCATGC 1030

Oy 127 aagacagagccacatgccctacacagatgctgtgtgacagaggtccagagataccttg 186
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 1031 AAGACAGGAGCCACATGCCCTACACAGATGCTGTGTGCACGAGGTCCAGAGATACCTTG 1090

Oy 187 acctctccccaccagctgccccatgcacgtgacacctgacattaaattcagaaactatc 246
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 1091 ACCTTCTCCCCACAGCCTGCCCATGCAGTGCACCTGTGACATTAAATTCAGAAACTATC 1150

Oy 247 tcattcccaagg 258
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 1151 TCATTCCCAAGG 1162

RESULT 13
E10631 standard; RNA: HUM; 1473 BP.
ID E10631
XX
AC E10631;
XX
SV E10631.1
XX
XX
DT 08-OCT-1997 (Rel. 52, Created)
DT 02-SEP-2000 (Rel. 65, Last updated, Version 2)
XX
DE Human cDNA encoding cytochrome P4502C9.
XX
KW JP 1996027196-A/5.
XX
OS Homo sapiens (human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
XX
```

```
[1]
RN 1-1473
RA Funae Y., Imaoka S., Matsuki Y., Hayashi K., Yabusaki Y.:
RT *ANTIBODY RECOGNIZING CYTOCHROME P450C9 ORIGINATED FROM MAN*.
RL Patent number JP1996027196-A/5, 30-JAN-1996.
RL SUMITOMO CHEM CO LTD.
XX
CC OS Homo sapiens (human)
CC PN JP 1996027196-A/5
CC PD 30-JAN-1996
CC PF 13-JUL-1994 JP 1994161551
CC PI FUNAE YOSHIHIKO, IMAOKA SUSUMU, MATSUKI YASUSHI, HAYASHI KOJI,
CC YABUSAKI YOSHIYASU
CC PC C07K16/18.C12N15/09.G01N33/53.G01N33/53//C12N1/19.C12N9/02,
CC (C12N1/19,
CC C12R1:865);
CC CC strandedness: Double;
CC CC topology: Linear;
CC CC Key Location/Qualifiers
CC FH Key
CC FT source 1..1473
CC FT /organism="Homo sapiens"
CC FT /tissue_type="liver"
CC FT CDS 1..1473
CC FT /product="cytochrome P450C9"
XX
FH Key Location/Qualifiers
FH FT source 1..1473
FH FT /db_xref="taxon:9606"
FH FT /organism="Homo sapiens"
XX
SQ Sequence 1473 BP; 413 A; 345 C; 319 G; 396 T; 0 other;

Query Match 58.9%; Score 190.4; DB 22; Length 1473;
Best Local Similarity 99.5%; Pred. No. 2.1e-45;
Matches 191; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 67 cagctaaagtcaggagagattgaacgtgtgattggcagaaaccggagccctgcatgc 126
DB 959 CAGCTAAAGTCCAGGAGAGATTGAACGTGTGATTGGCAGAAACCGGAGCCCTGCAATG 1018
QY 127 aagacaggagccacatgccctacacagatgctgtgtgcacagaggtccagagatacttg 186
DB 1019 AAGACAGGAGCCACATGCCCTACACAGATGCTGTGTGCACGAGGTCCAGAGATACATTG 1078
QY 187 accttctcccacagcctgccctacacagatgctgtgtgcacagaggtccagagatacttg 246
DB 1079 ACCTTCTCCCACAGCCTGCCCTACACAGATGCTGTGTGCACGAGGTCCAGAGATACATTG 1138
QY 247 tcattcccaagg 258
DB 1139 TCATTCCCAAGG 1150

RESULT 14
E10E53
ID E10853 standard; RNA; HUM; 1473 BP.
XX
AC E10853;
XX
SV E10853.1
XX
DT 07-OCT-1997 (Rel. 52, Created)
DT 02-SEP-2000 (Rel. 65, Last updated, Version 2)
XX
DE cDNA encoding human cytochrome P450.
XX
KW JP 1996056695-A/2.
XX
OS Homo sapiens (human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;
```

```
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
XX
RN 1-1473
RA Hayashi K., Sakaki T., Yabusaki Y., Komai K., Kaneko H., Nakatsuka I.:
RT *METHOD FOR EVALUATING SAFETY*.
RL Patent number JP1996056695-A/2, 05-MAR-1996.
RL SUMITOMO CHEM CO LTD.
XX
CC OS Homo sapiens (human)
CC PN JP 1996056695-A/2
CC PD 05-MAR-1996
CC PF 15-JUL-1994 JP 1994164184
CC PC 20-JUL-1993 JP 93P 201120, 30-JUL-1993 JP 93P 208279,
CC PR 17-JUN-1994 JP 94P 136053
CC PI HAYASHI KOJI, SAKAKI TOSHIYUKI, YABUSAKI YOSHIYASU,
CC KOMAI KOICHIRO,
CC KANEKO HIDEO, NAKATSUKA IWAO
CC PC C12Q1/02.C12M1/34.C12Q1/26;
CC CC strandedness: Double;
CC CC topology: Linear;
CC CC hypothetical: No;
CC CC anti-sense: No;
CC CC Key Location/Qualifiers
CC FH Key
CC FT source 1..1473
CC FT /organism="Homo sapiens"
CC FT CDS 1..1473
CC FT /product="human cytochrome P450 2C9"
XX
FH Key Location/Qualifiers
FH FT source 1..1473
FH FT /db_xref="taxon:9606"
FH FT /organism="Homo sapiens"
XX
SQ Sequence 1473 BP; 413 A; 345 C; 319 G; 396 T; 0 other;

Query Match 58.9%; Score 190.4; DB 22; Length 1473;
Best Local Similarity 99.5%; Pred. No. 2.1e-45;
Matches 191; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 67 cagctaaagtcaggagagattgaacgtgtgattggcagaaaccggagccctgcatgc 126
DB 959 CAGCTAAAGTCCAGGAGAGATTGAACGTGTGATTGGCAGAAACCGGAGCCCTGCAATG 1018
QY 127 aagacaggagccacatgccctacacagatgctgtgtgcacagaggtccagagatacttg 186
DB 1019 AAGACAGGAGCCACATGCCCTACACAGATGCTGTGTGCACGAGGTCCAGAGATACATTG 1078
QY 187 accttctcccacagcctgccctacacagatgctgtgtgcacagaggtccagagatacttg 246
DB 1079 ACCTTCTCCCACAGCCTGCCCTACACAGATGCTGTGTGCACGAGGTCCAGAGATACATTG 1138
QY 247 tcattcccaagg 258
DB 1139 TCATTCCCAAGG 1150

RESULT 15
S46963
LOCUS S46963 1814 bp mRNA PRI 05-AUG-1999
DEFINITION putative CYP2C9 gene [human, liver, mRNA, 1814 nt].
ACCESSION S46963
VERSION S46963.1 GI:258514
KEYWORDS human liver.
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1814)
AUTHORS Ohguya,S., Komori,M., Ohi,H., Shiramatsu,K., Shinriki,N. and
```

Kamatani, T.
Six-base deletion occurring in messages of human cytochrome P-450
in the CYP2C subfamily results in reduction of tolbutamide
hydroxylase activity
Biochem. Int. 27 (6), 1073-1081 (1992)
93075249
GenBank staff at the National Library of Medicine created this
entry [NCBI gibbsg 117179] from the original journal article.
This sequence comes from Fig. 1.
Author also gives sequence for another cytochrome P-450 that
contains a 6-base deletion of nucleotides 783-788.
Location/Qualifiers
1. .1814
/organism="Homo sapiens"
/db_xref="taxon:9606"
1. .1814
/gene="putative CYP2C9 gene"
3. .1436
/gene="putative CYP2C9 gene"
/note="This sequence comes from Fig. 1"
/codon_start=1
/product="cytochrome P-450"
/protein_id="AAB23864.2"
/db_xref="GI:5705937"
/translation="MLLSLWRQSGRGKLPPTPLPVGNIQLQIKDISKLTNL
SKVYGPVTLFGLKPIVVLHGVEAVKEALIDLGEEFSGRGIFPLAERANRGFGIVFS
NGKWKVEIRFSLMTLRNFGMKRSIEDVOEAEARCLVEELRKTASPCDPIFILGCA
PCNVICSIIFHKRFDYKQDFNLMLKLNENIKILSPWQICNNFPIIDYFPCTHN
KLLKNVAFMKSYILEKVKHOESMDMNPQDFIDCFMKMEKEKHNPQSEFTIESLEN
TAVDLFGAGTETTSITLRYALLLLKHPEVTAKVOEETERVIGNRSPCWDKSHMPY
TDVYVHQRTIDILPTSLPHAVICDIFRNYLIPKGTIIILISLTVLHDKNEFPNPE
MFDPHHFLOEGNFKSKSYFMPFSGAKRICVGEALAGMELFLFSLTILQNFNLKSLVD
PKNLDITPVVNGFASVPFPYQCFIPV"
BASE COUNT 517 a 413 c 369 g 515 t
ORIGIN

Query Match 58.9%; Score 190.4; DB 9; Length 1814;
Best Local Similarity 99.5%; Pred. No. 2.2e-45;
Matches 191; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 67 cagctaaagtcaggagagattgaacgtgtgattggcagaaacggagccctgcatgc 126
|||||
Db 922 CAGCTAAAGTCAGGAAGAGATTGACGTGTGATGGCAGAAACCGGACCCCTGCATGC 981
QY 127 aagacaggagccacatgcccctacacagatgctgtgtgacagaggtccagagataccttg 186
|||||
Db 982 AAGACAGGAGCCACATGCCCTACACAGATGCTGTGTGTCACGAGGTCCAGAGATACATTG 1041
QY 187 acctctccccaccagcctgccctcagtgacgtgacattaaattcagaaactatc 246
|||||
Db 1042 ACCTTCTCCCAAGGCTGCCCTGACCTGACCTGTGACATTAAATTGAGAAACTATC 1101
QY 247 tcattcccaagg 258
|||||
Db 1102 TCATTCCCAAGG 1113

Search completed: April 19, 2002, 09:34:57
Job time: 7997 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 19, 2002, 08:10:39 ; Search time 130.44 Seconds
(without alignments)
560.813 Million cell updates/sec

Title: us-09-763-292-3
Perfect score: 333
Sequences: 1 cccctgaattgctaacacaa.....aaattcatgatacttttt 333

Scoring table: IDENTITY.MSC
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0

Maximum Match 100

Listing first 45 summaries

Database : Issued Patents NA:
1: /cgn2.6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2.6/ptodata/2/ina/5A_COMB.seq.*
3: /cgn2.6/ptodata/2/ina/5A_COMB.seq.*
4: /cgn2.6/ptodata/2/ina/5A_COMB.seq.*
5: /cgn2.6/ptodata/2/ina/PTGUS_COMB.seq.*
6: /cgn2.6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	192	59.4	1854	1	US-08-201-118-4
2	192	59.4	1854	2	US-08-201-118-4
3	192	59.4	1854	3	US-08-201-118-4
4	190.4	58.9	1852	1	US-08-218-821B-10
5	190.4	58.9	1852	2	US-08-218-821B-10
6	190.4	58.9	1852	3	US-08-218-821B-10
7	188.8	58.5	1852	1	US-08-194-981E-3
8	188.8	58.5	1852	2	US-08-194-981E-3
9	188.8	58.5	1852	3	US-08-194-981E-3
10	180.6	55.9	1892	1	US-08-238-821B-14
11	180.6	55.9	1892	2	US-08-238-821B-14
12	177.6	55.0	1746	1	US-08-238-821B-2
13	177.6	55.0	1746	2	US-08-238-821B-2
14	177.6	55.0	1746	3	US-08-238-821B-2
15	168	52.0	2009	1	US-08-201-118-6
16	168	52.0	2009	2	US-08-218-821B-6
17	168	52.0	2009	3	US-08-218-821B-6
18	168	52.0	2258	1	US-08-201-118-12
19	168	52.0	2258	2	US-08-218-821B-12
20	168	52.0	2258	3	US-08-218-821B-12
21	153.6	47.6	1829	1	US-08-201-118-8
22	153.6	47.6	1829	2	US-08-238-821B-8
23	153.6	47.6	1829	3	US-08-238-821B-8
24	150.4	46.1	1829	1	US-08-194-981E-1
25	150.4	46.1	1829	2	US-08-194-981E-1
26	91	28.2	8779	2	US-08-750-703-1
27	83.4	25.8	1737	2	US-08-750-703-2

Sequence 5, Appl
Sequence 3, Appl
Sequence 1, Appl
Sequence 17, Appl
Sequence 14, Appl
Sequence 23, Appl
Sequence 16, Appl
Sequence 19, Appl
Sequence 13, Appl
Sequence 20, Appl
Sequence 21, Appl
Sequence 18, Appl
Sequence 15, Appl
Patent No. 5508199
Sequence 1, Appl
Sequence 1, Appl

ALIGNMENTS

US-08-201-118-4
: Sequence 4, Application US/08201118
: Patent No. 5786191
: GENERAL INFORMATION:
: APPLICANT: ROBERTSON, Joyce A
: APPLICANT: ROKES-SPARKS, Marjorie
: TITLE OF INVENTION: CLONING AND EXPRESSION OF COMPLEMENTARY
: TITLE OF INVENTION: DNAS FOR MULTIPLE MEMBERS OF THE HUMAN CYTOCHROME P450 2C
: NUMBER OF SEQUENCES: 44
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Townsend and Townsend Kourie and Crew
: STREET: 379 Lytton Avenue
: CITY: Palo Alto
: STATE: California
: COUNTRY: US
: ZIP: 94301
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.25
: CURRENT APPLICATION NUMBER: US/08201118
: APPLICATION NUMBER: 05/08201118
: CLASSIFICATION: 22 FEB 1994
: CLASSIFICATION: 433
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/864,962
: FILING DATE: 09-APR-1992
: ATTORNEY/AGENT INFORMATION:
: NAME: Liebeschuetz, Joe
: REGISTRATION NUMBER: 37,505
: REFERENCE/DOCKET NUMBER: 15280-192-1
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 329-6400
: TELEFAX: (415) 329-6422
: INFORMATION FOR SEQ ID NO. 4:
: SEQUENCE CHARACTERISTICS:
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
US-08-201-118-4

Query Match 59.4%, Score 192; DB 1; Length 1854;
Best Local Similarity 100.0%; Pred. No. 2 (e-5);
Matches 192; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

: MOLECULE TYPE: CDNA
PCT-US95-05744-10

	Query Match	58.9%	Score 190.4	DB 5	Length 1852
	Best Local Similarity	99.5%	Prod. No. 6.8e-51		
	Matches 191	Conservative	0	Mismatches 1	Indels 0
					Gaps 0
67	ccgcctaaagctccaggaagattgaacgctgtgattggcgaagaccgagccctgcacgc	126			
969	caccttaagctccaggaagattgaacgctgtgattggcgaagaccgagccctgcacgc	1028			
127	aagcagcagagccacagccctaccacagctcgtgctgcagcagaggtccagagatacctg	186			
1029	aagcacagagccacacatgcaccttacacagatcgtgtgtgctgcagcaggtctccagagatacattg	1088			
187	acctctctcccaccacagcctgcgcccaagcagctgacccctgacataaattccgaagaactatc	246			
1089	acctctctcccaccacagcctgcgcccaagcagctgacccctgacataaattccgaagaactatc	1148			
247	tcattcccaagg	258			
1149	tcattcccaagg	1160			

US-08-194-981E-4

Sequence 4, Application US/08194981E

Patent No. 5886157

GENERAL INFORMATION:

APPLICANT: GUENGERICH, F. Peter

APPLICANT: GUO, Zuyu

APPLICANT: KIM, JUNGSEUNG

APPLICANT: GILLIAM, Elizabeth M. J.

TITLE OF INVENTION: EXPRESSION AND PURIFICATION OF

TITLE OF INVENTION: HUMAN

TITLE OF INVENTION: CYTOCHROME P450

NUMBER OF SEQUENCES: 68

CORRESPONDENCE ADDRESS:

ADDRESSEE: NEEDLE & ROSENBERG, P.C.

STREET: Suite 1200, 127 Peachtree Street, NE

CITY: Atlanta

STATE: Georgia

COUNTRY: USA

ZIP: 30303-1811

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

RELEASE DATE: Patent Office Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/194,981E

FILING DATE: February 10, 1994

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Elizabeth Selby

REGISTRATION NUMBER: 38,298

REFERENCE/DOCKET NUMBER: 22000.0022

TELEPHONE: (404) 688-9810

TELEPHONE: (404) 688-0770

TELEFAX: (404) 688-9880

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 1419 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

ORGANISM: HUMAN

ANTI-SENSE: NO

US-08-194-981E-4

Query Match 58.5%; Score 188.8; DB 2; Length 1419;
Best Local Similarity 99.0%; Pred. No. 2e-50;
Matches 190; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 67 cagctaaagtcacagagattggaacgtgtagtgacagaaacccgagccctcgcacg 126
Db 905 CAGCTAAAGTCACGAGAGATTGGAACGTGTGTTGGCAGAAACCGAGCCCTCGCATGC 964
Oy 127 agagcaggagccacatgcctctacacagatgctgtgtgacagaggttcagagataccttg 186
Db 965 AGACAGAGGCGACATGCCCTACAGATGCTGTGTGTGACAGAGGTTCAGAGATGCATTC 1024
Oy 187 acctctccccacagctgccccacagatgcagtgacctgtgcatttaattcagaactatc 246
Db 1025 ACCTTCTCCACAGCTGCCCTACAGATGCTGTGTGTGACAGAGGTTCAGAGATGCATTC 1084
Oy 247 tcaattcccaagg 258
Db 1085 TCATTCCCAAG 1096

RESULT 8
US-08-194-981E-3
: Sequence 3, Application US/08194981E
: Patent No. 5786191
: GENERAL INFORMATION:
: APPLICANT: GRENKE, F. Peter
: APPLICANT: GRENKE, F. Peter
: APPLICANT: SANDHU, Punam
: TITLE OF INVENTION: EXPRESSION AND PURIFICATION OF
: TITLE OF INVENTION: HUMAN
: TITLE OF INVENTION: CYTOCHROME P450
: NUMBER OF SEQUENCES: 68
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: NEEDLE & ROSENBERG, P.C.
: STREET: Suite 1200, 127 Peachtree Street, N2
: CITY: Atlanta
: STATE: Georgia
: COUNTRY: USA
: ZIP: 30303-1811
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent in Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/194,981E
: FILING DATE: February 10, 1994
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Elizabeth Selby
: REGISTRATION NUMBER: 38,298
: REFERENCE/DOCKET NUMBER: 22000, 0022
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (404) 688-0770
: TELEFAX: (404) 688-8880
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1591 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: HYPOTHETICAL: NO
: AMTI-SENSE: NO
US-08-194-981E-3

Query Match 58.5%; Score 188.8; DB 2; Length 1591;
Best Local Similarity 99.0%; Pred. No. 2.1e-50;
Matches 190; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 67 cagctaaagtcacagagattggaacgtgtagtgacagaaacccgagccctcgcacg 126
Db 959 CAGCTAAAGTCACGAGAGATTGGAACGTGTGTTGGCAGAAACCGAGCCCTCGCATGC 1018
Oy 127 agagcaggagccacatgcctctacacagatgctgtgtgacagaggttcagagataccttg 186
Db 1019 AGACAGAGGCGACATGCCCTACAGATGCTGTGTGTGACAGAGGTTCAGAGATGCATTC 1078
Oy 187 acctctccccacagctgccccacagatgcagtgacctgtgcatttaattcagaactatc 246
Db 1079 ACCTTCTCCACAGCTGCCCTACAGATGCTGTGTGTGACAGAGGTTCAGAGATGCATTC 1138
Oy 247 tcaattcccaagg 258
Db 1139 TCATTCCCAAG 1150

RESULT 9
US-08-201-118-14
: Sequence 14, Application US/08201118
: Patent No. 5786191
: GENERAL INFORMATION:
: APPLICANT: GOLDSTEIN, Joyce A.
: APPLICANT: ROMKES-SPARKS, Marjorie
: TITLE OF INVENTION: CLONING AND EXPRESSION OF COMPLEMENTARY
: TITLE OF INVENTION: DNAs FOR MULTIPLE MEMBERS OF THE HUMAN CYTOCHROME P450 2C
: TITLE OF INVENTION: SUBFAMILY
: NUMBER OF SEQUENCES: 14
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Townsend and Townsend Kourile and Crew
: STREET: 379 Lytton Avenue
: CITY: Palo Alto
: STATE: California
: COUNTRY: US
: ZIP: 94301
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent in Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/201,118
: FILING DATE: 22-FEB-1994
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/864,962
: FILING DATE: 09-APR-1992
: ATTORNEY/AGENT INFORMATION:
: NAME: Goldstein, Joyce A.
: REGISTRATION NUMBER: 37,505
: REFERENCE/DOCKET NUMBER: 15280-192-1
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 326-2400
: TELEFAX: (415) 326-2422
: INFORMATION FOR SEQ ID NO: 14:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1892 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: HYPOTHETICAL: YES
US-08-201-118-14

Query Match 55.9%; Score 180.6; DB 1; Length 1892;
Best Local Similarity 95.3%; Pred. No. 8.7e-48;
Matches 183; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Oy 67 cagctaaagtcacagagattggaacgtgtagtgacagaaacccgagccctcgcacg 126
Db 1001 CAGCTAAAGTCACGAGAGATTGGAACGTGTGTTGGCAGAAACCGAGCCCTCGCATGC 1060

07 127 aagacagagccacatgccctacacagatgctgtgtgacagaggtccagagatacttg 186
 Db 1061 AGACAGAGCCACATGCCCTACACAGATGCTGTGTGACAGAGGTCCAGAGATACTTG 1120
 07 187 acctctcccccacagcctccacagcctgagcctgtgacattaaattccgaaaactatc 245
 Db 1121 ACTCTCTCCCCACAGCCTCCACAGCCTGAGCCTGTGACATTAAATTCCGAAACTACC 1180
 07 247 tcaattcccaagg 258
 Db 1181 TCATNCCCAAGG 1192

RESULT 10
 US-08-238-821B-14
 : Sequence 14, Application US/08238821B
 : Patent No. 5912120
 : GENERAL INFORMATION:
 : APPLICANT: GOLDSTEIN, Joyce A.
 : APPLICANT: ROMEK-SPARKS, Marjorie
 : APPLICANT: DE MORALS, Sonia M.F.
 : TITLE OF INVENTION: CLONING, EXPRESSION AND DIAGNOSIS OF HUMAN
 : TITLE OF INVENTION: CYTOCHROME P450 2C19: THE PRINCIPAL DETERMINANT OF S-
 : TITLE OF INVENTION: MEPHYNTOLIN METABOLISM
 : NUMBER OF SEQUENCES: 61
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Townsend and Townsend and Crew LLP
 : STREET: Two Embarcadero Center, 8th Floor
 : CITY: San Francisco
 : STATE: California
 : COUNTRY: US
 : ZIP: 94111
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: Patent in Release #1.0, Version #1.25
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/238.821B
 : FILING DATE: 06-MAY-1994
 : CLASSIFICATION: 435
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: US 08/201,118
 : FILING DATE: 27-FEB-1994
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: 435
 : FILING DATE: 06-MAY-1994
 : CLASSIFICATION: 435
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: US 07/864,962
 : FILING DATE: 09-APR-1992
 : ATTORNEY/AGENT INFORMATION:
 : NAME: Liebeschuetz, Joe
 : REGISTRATION NUMBER: 37,505
 : REFERENCE/DOCKET NUMBER: 15280-1921100S
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: (650) 326-2400
 : TELEFAX: (650) 326-2422
 : INFORMATION FOR PUBLICATION:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 1892 base pairs
 : TYPE: nucleic acid
 : STRANDEDNESS: single
 : TOPOLOGY: linear
 : MOLECULE TYPE: DNA (genomic)
 : FEATURE:
 : NAME/KEY: Region
 : LOCATION: 1..41
 : OTHER INFORMATION:
 : OTHER INFORMATION: /note= *Corresponds to positions -41 to -1
 : OTHER INFORMATION: for 2c of figure 2.

Query Match 55.9%; Score 180.6; DB 2; Length 1892;
 Best Local Similarity 95.3%; Pred. No. 8.7e-48;

Matches 183: Conservative 0: Mismatches 9: Indels 0: Gaps 0:
 07 67 cagctaaagtcacagagagattgaacctgtatgttgacgaaacaggagccctgcatgc 126
 Db 1001 CAGCTAAAGTCCAGGAAGAGATTGACAGCTGTATTTGCGAAGAACCGAGCCCTGCTCATGC 1060
 07 127 aagacagagccacatgccctacacagatgctgtgtgacagaggtccagagatacttg 186
 Db 1061 AGACAGAGCCACATGCCCTACACAGATGCTGTGTGACAGAGGTCCAGAGATACTTG 1120
 07 187 acctctcccccacagcctccacagcctgagcctgtgacattaaattccgaaaactatc 245
 Db 1121 ACTCTCTCCCCACAGCCTCCACAGCCTGAGCCTGTGACATTAAATTCCGAAACTACC 1180
 07 247 tcaattcccaagg 258
 Db 1181 TCATNCCCAAGG 1192

RESULT 11
 PCT-US95-05744-14
 : Sequence 14, Application PCT/US9505744
 : GENERAL INFORMATION:
 : APPLICANT: GOLDSTEIN, Joyce A.
 : APPLICANT: ROMEK-SPARKS, Marjorie
 : APPLICANT: DE MORALS, Sonia M.F.
 : TITLE OF INVENTION: CLONING, EXPRESSION AND DIAGNOSIS OF HUMAN
 : TITLE OF INVENTION: CYTOCHROME P450 2C19: THE PRINCIPAL DETERMINANT
 : NUMBER OF SEQUENCES: 61
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Townsend and Townsend Kourile and Crew
 : STREET: 379 Fulton Avenue
 : CITY: Palo Alto
 : STATE: California
 : COUNTRY: US
 : ZIP: 94301
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: Patent in Release #1.0, Version #1.25
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: PCT/US95/05744
 : FILING DATE:
 : CLASSIFICATION:
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: US 08/238,821
 : FILING DATE: 06-MAY-1994
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: US 08/201,118
 : FILING DATE: 22-FEB-1994
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: US 07/864,962
 : FILING DATE: 09-APR-1992
 : ATTORNEY/AGENT INFORMATION:
 : NAME: Dow, Karen B.
 : REGISTRATION NUMBER: 29,684
 : REFERENCE/DOCKET NUMBER: 15280-192-1-1
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: (415) 326-2400
 : TELEFAX: (415) 326-2422
 : INFORMATION FOR PUBLICATION:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 1892 base pairs
 : TYPE: nucleic acid
 : STRANDEDNESS: single
 : TOPOLOGY: linear
 : MOLECULE TYPE: DNA (genomic)
 : HYPOTHETICAL: YES
 : PCT-US95-05744-14

Query Match 55.9%; Score 180.6; DB 5; Length 1892;
Best Local Similarity 95.3%; Pred. No. 7.9e-48;
Matches 183; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

07 67 cagctaaatgctcaggaagattgaacgtgtgattgacgaggaacccctgcacgc 126
DB 1001 CAGCTAAATGCTCAGGAGATTGAAGTGTGATTGCGAGAAACCGAGCCCTGCATGC 1060

07 127 aagacagagccacatccctacacagatgctgtgtgacgaggtccagagataccttg 186
DB 1061 AAGACAGAGCCACATCCCTACACAGATGCTGTGTGCGACGAGGTCCAGAGATACATGC 1120

07 187 acctctccccacagctgccccatgacgtgacgtgacatttaattcagaatactc 246
DB 1121 ACCTCTCCCCACAGCTGCCCATGACGTTGATGCTGATTAATTTCAGAACTACC 1180

07 247 tcatcccaagg 258
DB 1181 TCATCCCAAGG 1192

RESULT 12
US-08-201-118-2
Sequence 2, Application 05/08201118
Patent No. 5786131
GENERAL INFORMATION:
APPLICANT: GOLDSTEIN, Joyce A.
APPLICANT: ROMKES-SPARKS, Marjorie
TITLE OF INVENTION: CLONING AND EXPRESSION OF COMPLEMENTARY
TITLE OF INVENTION: DNAs FOR MULTIPLE MEMBERS OF THE HUMAN CYTOCHROME P450 2C
NUMBER OF SEQUENCES: 44
SUBFAMILY
CORRESPONDENCE ADDRESS:
ADDRESS: Townsend and Townsend Kourie and Crew
STREET: 379 Lytton Avenue
CITY: Palo Alto
STATE: California
COUNTRY: US
ZIP: 94301
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/201.118
FILING DATE: 22-FEB-1994
CLASSIFICATION: 435
PRIORITY INFORMATION:
APPLICATION NUMBER: US 07/864,962
FILING DATE: 09-APR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Liebeschuetz, Joe
REGISTRATION NUMBER: 37,505
REFERENCE/DOCKET NUMBER: 15280-192-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 326-2400
TELEFAX: (415) 326-2422
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1746 base pairs
TYPE: nucleic acid
STRANDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: Region
LOCATION: 1..5
OTHER INFORMATION: /note= "Corresponds to positions 15 to 1
US-08-201-118-2

Query Match 55.0%; Score 177.6; DB 1; Length 1746;
Best Local Similarity 95.3%; Pred. No. 7.9e-47;
Matches 183; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

07 67 cagctaaatgctcaggaagattgaacgtgtgattgacgaggaacccctgcacgc 126

DB 964 CAGCTAAATGCTCAGGAGATTGAACGTGTGATTGCGAGAAACCGAGCCCTGCATGC 1023

07 127 aagacagagccacatccctacacagatgctgtgtgacgaggtccagagataccttg 186

DB 1024 AAGACAGAGCCACATCCCTACACAGATGCTGTGTGCGACGAGGTCCAGAGATACATGC 1083

07 187 acctctccccacagctgccccatgacgtgacgtgacatttaattcagaatactc 246

DB 1084 ACCTCTCCCCACAGCTGCCCATGACGTTGATGCTGATTAATTTCAGAACTACC 1143

07 247 tcatcccaagg 258

DB 1144 TCATCCCAAGG 1155

RESULT 13
US-08-238-821B-2
Sequence 2, Application US/08238821B
Patent No. 5912120
GENERAL INFORMATION:
APPLICANT: GOLDSTEIN, Joyce A.
APPLICANT: ROMKES-SPARKS, Marjorie
TITLE OF INVENTION: CLONING, EXPRESSION AND DIAGNOSIS OF HUMAN
TITLE OF INVENTION: CYTOCHROME P450 2C19: THE PRINCIPAL DETERMINANT OF S-
NUMBER OF SEQUENCES: 61
CORRESPONDENCE ADDRESS:
ADDRESS: Townsend and Townsend and Crew LLP
STREET: 800 California Center, 8th Floor
CITY: San Francisco
STATE: California
COUNTRY: US
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/238.821B
FILING DATE: 06-MAY-1994
CLASSIFICATION: 435
PRIORITY INFORMATION:
APPLICATION NUMBER: US 08/201.118
FILING DATE: 22-FEB-1994
CLASSIFICATION: 435
PRIORITY INFORMATION:
APPLICATION NUMBER: US 07/864,962
FILING DATE: 09-APR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Liebeschuetz, Joe
REGISTRATION NUMBER: 37,505
REFERENCE/DOCKET NUMBER: 15280-1921100S
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 326-2400
TELEFAX: (650) 326-2422
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1746 base pairs
TYPE: nucleic acid
STRANDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: Region
LOCATION: 1..5
OTHER INFORMATION: /note= "Corresponds to positions 15 to 1
US-08-238-821B-2

Query Match 55.0%; Score 177.6; DB 2; Length 1746;
 Best Local Similarity 95.3%; Pred. No. 7.5e-47;
 Matches 183; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 67 cagctaaagtcacgaagagattgaacgtgtgtagtcagaaacccgagccctcgatgc 126
 DB 964 CAGCTAAAGTCCAGAGAGATTGAACGTGTGTTGGCAGAACCCGAGCCCTCGATGC 1023

QY 127 aagcaggagccatcgctccctacacagatgctgtgtagtcagaaacccgagccctcgatgc 186
 DB 1024 AGACAGAGGCGCCACATGCTTACACAGATGCTGTGTGTCAGCAGGTCCAGAGATACATCG 1083

QY 187 accttctcccccagcctgcccctacacagatgctgtgtagtcagaaacccgagccctcgatgc 246
 DB 1084 ACTTCATCCACAGCCTGCCCCACAGATGCTGTGTGTCAGCAGGTCCAGAGATACATCG 1143

QY 247 tcattcccaagg 258
 DB 1144 TCATTCCCAAGG 1155

RESULT 14
 PCT-US95-05744-2
 : Sequence 2, Application PC/TUS9505744
 : GENERAL INFORMATION:
 : APPLICANT: GOLDSTEIN, Joyce A.
 : APPLICANT: ROMKES-SPARKS, Marjorie
 : APPLICANT: DE MORAIS, Sonia M.F.
 : TITLE OF INVENTION: CLONING, EXPRESSION AND DIAGNOSIS OF HUMAN
 : TITLE OF INVENTION: CYTOCHROME P450 2C19: THE PRINCIPAL DETERMINANT
 : TITLE OF INVENTION: OF S-MEPHITOXIN METABOLISM
 : NUMBER OF SEQUENCES: 61
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Townsend and Townsend Kourie and Crew
 : STREET: 379 Lytton Avenue
 : CITY: Palo Alto
 : STATE: California
 : COUNTRY: US
 : ZIP: 94301
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : OPERATING SYSTEM: IBM PC compatible
 : SOFTWARE: Patent In Release #1.0, Version #1.25
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: PCT/US95/05744
 : FILING DATE:
 : CLASSIFICATION:
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: US 08/238,821
 : FILING DATE: 06-MAY-1994
 : APPLICATION DATA: US 08/201,118
 : FILING DATE: 27-FEB-1994
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: US 07/864,962
 : FILING DATE: 09-APR-1992
 : ATTORNEY/AGENT INFORMATION:
 : NAME: Dow, Karen B.
 : REGISTRATION NUMBER: 29,684
 : REFERENCE/DOCKET NUMBER: 15280-192-1-1
 : TELEPHONE: (415) 326-2400
 : TELEFAX: (415) 326-2400
 : INFORMATION FOR SEQ ID NO. 2:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 1746 base pairs
 : TYPE: nucleic acid
 : STRANDEDNESS: single
 : TOPOLOGY: linear
 : MOLECULE TYPE: cDNA
 : PCT-US95-05744-2

Query Match 55.0%; Score 177.6; DB 5; Length 1746;
 Best Local Similarity 95.3%; Pred. No. 7.5e-47;
 Matches 183; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 67 cagctaaagtcacgaagagattgaacgtgtgtagtcagaaacccgagccctcgatgc 126
 DB 964 CAGCTAAAGTCCAGAGAGATTGAACGTGTGTTGGCAGAACCCGAGCCCTCGATGC 1023

QY 127 aagcaggagccatcgctccctacacagatgctgtgtagtcagaaacccgagccctcgatgc 186
 DB 1024 AGACAGAGGCGCCACATGCTTACACAGATGCTGTGTGTCAGCAGGTCCAGAGATACATCG 1083

QY 187 accttctcccccagcctgcccctacacagatgctgtgtagtcagaaacccgagccctcgatgc 246
 DB 1084 ACTTCATCCACAGCCTGCCCCACAGATGCTGTGTGTCAGCAGGTCCAGAGATACATCG 1143

QY 247 tcattcccaagg 258
 DB 1144 TCATTCCCAAGG 1155

RESULT 15
 US-08-201-118-6
 : Sequence 6, Application US/08201118
 : Patent No. 5786191
 : GENERAL INFORMATION:
 : APPLICANT: GOLDSTEIN, Joyce A.
 : APPLICANT: ROMKES-SPARKS, Marjorie
 : TITLE OF INVENTION: CLONING AND EXPRESSION OF COMPLEMENTARY
 : TITLE OF INVENTION: DNAs FOR MULTIPLE MEMBERS OF THE HUMAN CYTOCHROME P450 2C
 : NUMBER OF SEQUENCES: 4
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Townsend and Townsend Kourie and Crew
 : STREET: 379 Lytton Avenue
 : CITY: Palo Alto
 : STATE: California
 : COUNTRY: US
 : ZIP: 94301
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : OPERATING SYSTEM: IBM PC compatible
 : SOFTWARE: Patent In Release #1.0, Version #1.25
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/201,118
 : FILING DATE: 22-FEB-1994
 : CLASSIFICATION: 435
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: US 07/864,962
 : FILING DATE: 09-APR-1992
 : ATTORNEY/AGENT INFORMATION:
 : NAME: Liebschuetz, Joe
 : REGISTRATION NUMBER: 37,505
 : REFERENCE/DOCKET NUMBER: 15280-192-1
 : TELEPHONE: (415) 326-2400
 : TELEFAX: (415) 326-2422
 : INFORMATION FOR SEQ ID NO: 6:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 2009 base pairs
 : TYPE: nucleic acid
 : STRANDEDNESS: single
 : TOPOLOGY: linear
 : MOLECULE TYPE: cDNA
 : US-08-201-118-6

Query Match 52.0%; Score 168; DB 1; Length 2009;
 Best Local Similarity 92.3%; Pred. No. 8.7e-44;
 Matches 177; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

07 67 cagctaaagtcacagagagattgaacgtgtgatigggcagaaacccggagcccccctgcatgc 126
Db 1158 CAGCTAAAGTCCAGGAGAGATTGAATGTGTAGTTGGCAGAAACCGGAGCCCTGTATGC 1217
07 127 aagacacagagccacatgccctacacagatgctgtggtgcaacaggtccagagagataccttg 186
Db 1218 AGGACAGAGTCAATGCTTAAACAGATGCTGTGTGTCACAGATCCAGAGATACATTG 1277
07 187 accttctcccaaccagcctgccccatgcagtcgaactgtgacatttaattcagagaactatc 246
Db 1278 ACCTCTCTCCCAACCACTGCCCCATGCCACCTGCAGTGCCTGTGTAAATTCAAAACTACC 1337
07 247 tcatcccaagg 258
Db 1338 TCTCCCAAGG 1349

Search completed: April 19, 2002, 08:10:43
Job time: 2943 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: **April 19, 2002, 08:15:28 ; Search time 277.57 Seconds**
(without alignments)

Title: US-09-763-292-3

Perfect score: 323
Sequence: 1 cccctgaattgctacaaa.....aaattcatagtatcattttt 323

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

```

minimum DB seq length: 0

```

Maximum DB seq length: 9
Maximum DB seq length: 200000000

Post-processing: Minimum Match 09

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_1101:*

```
1: /SIDS2/gcgdata/geneseq/geneseq.NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq.NA1981.DAT.*
```

2: /SIDS2/gcgdata/geneseq/geneseqn/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseqn/NA1982.DAT.*

4: /SIDS2/gcgdata/geneseq/geneseqn/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseqn/NA1984.DAT.*

3: /SID52/qcdata/geneseq/NA1984.DAT.*
6: /SID52/qcdata/geneseq/NA1985.DAT.*

```
7: /SID52/gcgdata/geneseq/geneseqn/NA1986.DAT.*
9: /csmc/geneseq/geneseqn/NA1986.DAT.*
```

8: /SIDS2/qcqdqdata/geneseq/geneseqn/NA1987.DAT.*
9: /SIDS2/qcqdqdata/geneseq/geneseqn/NA1988.DAT.*

10: /SID52/gcgcdata/geneseq/geneseq/NA1989.DAT:
11: /SID52/gcgcdata/geneseq/geneseq/NA1989.DAT:

11: /SIDS2/qcdata/geneseq/geneseq/NA1990.DAT:
12: /SIDS2/qcdata/geneseq/geneseq/NA1991.DAT:

13: /SIDS2/gcgcdata/geneseq/geneseqn/NA1992.DAT:

14: /SIBS2/gcgdata/geneseq/geneseq/NA1993.DAT:
15: /SIBS2/gcgdata/geneseq/geneseq/NA1994.DAT:

16: /SIDS2/qcgcdata/geneseq/geneseq/NAI1995.DAT:

17: /SID2/gcdata/geneseq/geneseq/NA1996.DAT:
18: /SID2/gcdata/geneseq/geneseq/NA1997.DAT:

```
18: /SIDS2/gcgdata/geneseq/geneseqn/NA1997.DAT:
19: /SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:
20:
```

```
20: /SIPS2/gcgcdata/geneseq/geneseqn/NA1999.DAT:
21: /SIPS2/gcgcdata/geneseq/geneseqn/NA2000.DAT:
```

22: /SIDS2/gcgsdata/geneseq/geneseqn/NA2001.DAT:

Pred. No. is the number of results predicted by chance to be

score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	276.8	85.7	8437	32	AA012242	Human Cytochrome P
2	223	69.0	1001	27	A0151162	Human CYP2C8 relat
3	192	59.4	1854	17	AA111378	Cytochrome P450 2C
4	192	59.4	1854	19	AAV41416	Human cytochrome P
5	190.4	58.9	1356	20	AA019916	Human cytochrome P
6	190.4	58.9	1473	16	AA087715	Human cytochrome P
7	190.4	58.9	1473	17	AA072818	Human cytochrome P
8	190.4	58.9	1473	17	AA017404	Human derived cyto
9	190.4	58.9	1818	11	AA003599	Human liver cyto
10	190.4	58.9	1845 ^a	20	AA119925	Mammalian cyto
11	190.4	58.9	1852	17	AA111381	Cytochrome P450 2C

Dd	1091	acctctccaccagcgcgcccacgtgacctgacattaaattcagaactatc
Oy	247	tcaatccaagg 258
Dd	1151	tcaatccaagg 1162
RESULT	5	
AAI19916		
D	AXI19916 standard; DNA; 1356 BP.	
C	AAI19916;	
AD	AXI19916;	
DT	11-JUN-1999 (first entry)	
DE	Bacterial and mammalian chimeric cytochrome P450 protein encoding DNA.	
KW	Bacterial; mammalian; cytochrome P450; chimeric; fusion protein;	
KX	Oxidises hydrocarbon; carbon-hydrogen bond; hydroxylating;	
KW	Bioremediation; environmental pollutant; ss.	
OS	Synthetic.	
PN	WO9908812-A1.	
PD	25-FEB-1999.	
PR	17-AUG-1998: 98WO-US16979.	
PR	20-AUG-1997: 97US-0056754.	
PPA	(UTRP) UNIV ROCHESTER.	
PI	Jones JP, Shimoji M;	
PT	WPI: 1999-190131/16.	
DR	P-PDB: AAY04126.	
PPT	New P450 fusion proteins - comprising a portion of a bacterial	
PPT	cytochrome P450 protein and a portion of a mammalian cytochrome P450	
PPT	protein	
CC	Claim 10; Page 5-6; Sipp; English.	
CCC	The present sequence encodes a fusion proteins comprising a portion of	
CCC	bacterial cytochrome P450 protein and also a portion of a mammalian or	
CCC	cytochrome P450 protein. The fusion protein can oxidise hydrocarbons or	
CCC	any compound having a carbon-hydrogen bond. The fusion protein can be used	
CCC	used for hydroxylating a compound to be oxidised. It can also be used i	
CCC	the bioremediation of an environmental pollutant. Since the fusion	
CCC	protein has been subjected to structural elucidation by x-ra	
CCC	crystallography for designing bioremediation, it can be readily	
CCC	expressed in soil bacteria to facilitate bioremediation.	
SSO	Sequence 1356 BP; 349 A; 386 C; 323 G; 298 T; 0 other;	

	Query Match	Best Local Similarity	58.9%	Score 190.4	DB 20:	Length 1356;
	Matches 191:	Conservative	0:	Mismatches 1:	Indels 0:	Gaps 0:
0y	67	cagctcaaaattccaggaagaattgaacgtgtgatttcgagaacacgcgccttgatgc	126			
0b	842	cagctcaaaattccaggaagaattgaacgtgtgttcgagaacacgcgccttgatgc	901			
0y	127	agacacgaggccacatgccttaccagatgtctgtgttcgacgaggttcgaagaatacttg	186			
0b	197	agacacgaggccacatgccttaccagatgtctgtgttcgacgaggttcgaagaatacttg	961			
0y	187	actctctccacacgcttcccacgacgtgcacatgacgtgcacatgacgtgcacatgac	246			
0b	962	actctctccacacgcttcccacgacgtgcacatgacgtgcacatgacgtgcacatgac	1021			

QY 247 tcaatccaag 258
 DB 1022 tcaatccaag 1033
 RESULT 6
 ID AQ087715 standard; cDNA; 1473 BP.
 AC AQ087715;
 DT 10-MAY-1995 (first entry)
 DE Human cytochrome P450 molecular species 2C9 cDNA.
 EN Human cytochrome P450; amplification; PCR; primer: expression vector;
 KW yeast NADPH-P450 reductase; safety; fusion protein; metabolite;
 EV carcinogen; mutagen; liver metabolism; ds.
 OS Homo sapiens.
 PI EP644267-A.
 IN 22-MAR-1995.
 PP 20-JUL-1994; 94EP-0111298.
 PR 21-JUL-1993; 93JP-0180246.
 PR 20-JUL-1993; 93JP-0201120.
 PR 30-JUL-1993; 93JP-0208279.
 XX (BATAV) HAYASHI K.
 PA (SUMO) SUMITOMO CHEM CO LTD.
 PI Hayashi K, Kaneko B, Komai K, Nakatsuka I, Sakaki T;
 PI Tanaka Y;
 DR NPI: 1995-116991/16.
 DR P-PSDB; AAR72361.
 XX Evaluation of safety of a chemical cpd. - using recombinant yeast
 PT expressing human cytochrome P450 and a yeast NADPH-P450 reductase
 PS Examples; Page 23-25; 124pp; English.
 CC The nucleotide sequence of the cDNA encoding the human cytochrome P450
 CC species 2C9 gene encodes a protein of 450 amino acids. The cDNA
 CC was amplified by PCR using primers AAT76925-8. The prod. was cloned into the
 CC vector p2C9 for the expression of the cytochrome P450 alone or p2C9R
 CC for co-expression with the yeast NADPH-P450 reductase.
 CC The vectors are used in a method for evaluating the safety of a chemical
 CC compound by reacting the chemical compound with recombinantly produced
 CC human cytochrome P450 molecular species 1A2 (AA087714), 2C9, 2E1
 CC (AA087716) or 3A4 (AA087717), or their auxiliary species and variants
 CC (AA087718-32), and yeast NADPH-P450 reductase, either as a fused protein
 CC or as a cell extract, and analysing the resulting metabolite to assess the
 CC safety of the chemical compound. The method is useful for determining
 CC whether the chemical compound or its metabolite will be converted into
 CC a carcinogenic or mutagenic form through metabolism in the liver.
 XX Sequence 1473 BP; 413 A; 345 C; 319 G; 396 T; 0 other:
 Query Match 58.9%; Score 190.4; DB 16; Length 1473;
 Best Local Similarity 99.5%; Pred. No. 2.2e-51;
 Matches 191; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 67 cagctaaagtcaggaagattgaacatgtatgttcgagaaacagagccctgacgc 126
 DB 959 cagctaaagtcaggaagattgaacatgtatgttcgagaaacagagccctgacgc 1018

QY 127 aagacagagagccacatgcctcacagatgctgtggtgacagaggtccagagataccttg 186
 DB 1019 aagacagagagccacatgcctcacagatgctgtggtgacagaggtccagagataccttg 1078
 QY 187 acctctctccacagagctgtggtccatgacgtgacatgaataaattcagaactatc 246
 DB 1079 acctctctccacagagctgtggtccatgacgtgacatgaataaattcagaactatc 1138
 QY 247 tcaatccaag 258
 DB 1139 tcaatccaag 1150
 RESULT 7
 ID AAT78381 standard; DNA; 1473 BP.
 AC AAT78381;
 DT 11-OCT-1996 (first entry)
 DE Human cytochrome P450 molecular species 2C9 gene.
 EN Human cytochrome P450; amplified; PCR; polymerase chain reaction; primer;
 KW liver; yeast; expression vector; NADPH-P450 reductase; ADH gene promoter;
 EV evaluation; safety; fusion protein; metabolite; detoxification;
 KW carcinogenic; ds.
 OS Homo sapiens.
 PI JP08056695-A.
 PN 05-MAR-1996.
 PD 15-JUL-1994; 94JP-0164184.
 PF 17-JUN-1994; 94JP-0136053.
 PR 20-JUL-1993; 93JP-0201120.
 PR 30-JUL-1993; 93JP-0208279.
 PA (SUMO) SUMITOMO CHEM CO LTD.
 XX NPI: 1996-182311/19.
 DR P-PSDB; AAR93168.
 XX Novel method for the evaluation of the safety of a cpd. - using a
 PT human cytochrome P450 and yeast NADPH reductase to determine whether
 PT the analyte cpd. is detoxified or metabolised to a carcinogen
 XX Example 1; Page 20-22; 74pp; Japanese.
 CC This is the nucleotide sequence of the human cytochrome P450 molecular
 CC species 2C9 gene which encodes a protein of 450 amino acids. The gene
 CC was amplified from a human liver derived cDNA library as 2 fragments of
 CC 0.9 and 0.6 kb using primers AAT76925-8. The prod. was cloned into the
 CC yeast expression vector pAAH5N to generate plasmid p2C9 for prodn. of
 CC the cytochrome only or into the vector pAHR to generate the plasmid
 CC p2C9R for co-prodn. with the yeast NADPH-P450 reductase. The sequence
 CC is placed under control of the yeast ADH gene promoter and terminator.
 CC The vectors are used in a method for evaluating the safety of a cpd. by
 CC co-transforming the recombinant plasmid with a yeast NADPH-P450 reductase
 CC moiety species 1A2 (AA087714), 2C9, 2E1 (AA087716), NADPH-P450 reductase
 CC or their variants (AAT78381-98) together with yeast NADPH-P450 reductase
 CC (either as a fused protein or as a cell extract) and analysing the
 CC resultant metabolite. The cpd. is considered "safe" if it is detoxified
 CC or not rendered carcinogenic or "unsafe" if it is not detoxified or is
 CC metabolised to a carcinogenic cpd.
 XX Sequence 1473 BP; 413 A; 345 C; 319 G; 396 T; 0 other:
 Query Match 58.9%; Score 190.4; DB 17; Length 1473;
 Best Local Similarity 99.5%; Pred. No. 2.2e-51;
 Matches 191; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 67 cagctaaagtcaggagagagattgaacgtgtgattggcagaacccgagccctgcatgc 126
 DB 959 cagctaaagtcaggagagagattgaacgtgtgattggcagaacccgagccctgcatgc 1018
 OY 127 aagacagagccacatgcctccacacagatgctgtgattgacagagagccctgcatgc 186
 DB 1019 aagacagagccacatgcctccacacagatgctgtgattgacagagagccctgcatgc 1078
 OY 187 acctctcccccacagcctgcccacatgcagctgacattgaattcagaactatc 246
 DB 1079 acctctcccccacagcctgcccacatgcagctgacattgaattcagaactatc 1138
 OY 247 tcatcccaagg 258
 DB 1139 tcatcccaagg 1150

RESULT 8

AAT17404
 ID AAT17404 standard; cDNA: 1473 BP.
 AC AAT17404;
 XX 01-JUN-1996 (first entry)
 DT Human derived cytochrome P4502C9 cDNA.
 DE Human derived cytochrome: P4502C9; commercial cDNA library; yeast;
 KW transfection; recombinant production; expression vector; mammal;
 KW immunisation; sensitisation; antibody; determination; detection;
 KW non-cross reactive; ds.
 OS Homo sapiens.

XX Key Location/Qualifiers
 FT CDS 1..1473
 FT /tag= a
 XX JP08027196-A.
 XX 30-JAN-1996.
 XX 13-JUL-1994; 94JP-0161551.
 XX 13-JUL-1994; 94JP-0161551.
 XX (SDMO) SUMITOMO CHEM CO LTD.
 XX WPI; 1996-136337/14.
 XX P-PSDB; AAR81465.
 XX Antibody recognising human derived cytochrome P4502C9 - allows
 PT specific detection of cytochrome P450 species in humans
 XX Example 1: Pages 11-13; 13pp: Japanese.

XX The present sequence encodes the human derived cytochrome (HDC)
 CC P4502C9 which was subcloned into a linearised HindIII-HindIII
 CC vector and transfected into a yeast strain. The resulting HDC
 CC were transfected with an expression vector and cultured in yeast.
 CC HDC was purified from the fraction, and used to immunise and
 CC sensitise a mammal. Blood was drawn from the mammal, and an
 CC anti-HDC antibody isolated. The antibody obtd. recognises HDC
 CC P4502C9, partic. at a serum dilution rate of 1:10000, and is
 CC substantially without cross reaction to other HDC P450 spp..

XX Sequence 1473 BP; 413 A; 345 C; 319 G; 396 T; 0 other;

Query Match 58.9%; Score 190.4; DB 17; Length 1473;
 Best Local Similarity 99.5%; Pred. No. 2.2e-51;
 Matches 191; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 67 cagctaaagtcaggagagattgaacgtgtgattggcagaacccgagccctgcatgc 126
 DB 959 cagctaaagtcaggagagattgaacgtgtgattggcagaacccgagccctgcatgc 1018
 OY 127 aagacagagccacatgcctccacacagatgctgtgattgacagagagccctgcatgc 186
 DB 1019 aagacagagccacatgcctccacacagatgctgtgattgacagagagccctgcatgc 1078
 OY 187 acctctcccccacagcctgcccacatgcagctgacattgaattcagaactatc 246
 DB 1079 acctctcccccacagcctgcccacatgcagctgacattgaattcagaactatc 1138
 OY 247 tcatcccaagg 258
 DB 1139 tcatcccaagg 1150

RESULT 9

AAO03599
 ID AAO03599 standard; DNA: 1818 BP.
 AC AAO03599;
 XX 03-SEP-1990 (first entry)
 DT Human liver cytochrome P-450 encoding gene.
 DE P-450; cytochrome; Saccharomyces cerevisiae;
 KW liver disorders; ds.
 OS Homo sapiens.

XX Key Location/Qualifiers
 FT CDS 1..1443
 FT /tag= a
 XX JP02072879-A.
 XX 13-MAR-1990.
 XX 09-SEP-1988; 88JP-0225955.
 XX 09-SEP-1988; 88JP-0225955.
 XX (AGEN) AGENCY OF IND SCI TECH.
 XX WPI; 1990-121045/16.
 XX P-PSDB; AAR04045.
 XX Recombinant plasmid, for yeast for liver treatment -
 PT comprises human liver cytochrome P-450MP gene obtd. from yeast.
 PT etc.
 XX Disclosure: Fig 4; 9pp: Japanese.

XX The P-450 produced from plasmid pHPA6 in a yeast expression system
 CC preferably Saccharomyces AD22, can be used to treat human liver
 CC disorders, oxidising various chemical substances.
 XX Sequence 1818 BP; 516 A; 412 C; 376 G; 514 T; 0 other;

Query Match 58.9%; Score 190.4; DB 11; Length 1818;
 Best Local Similarity 99.5%; Pred. No. 2.4e-51;
 Matches 191; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 67 cagctaaagtcaggagagattgaacgtgtgattggcagaacccgagccctgcatgc 126
 DB 932 cagctaaagtcaggagagattgaacgtgtgattggcagaacccgagccctgcatgc 991
 OY 127 aagacagagccacatgcctccacacagatgctgtgattgacagagagccctgcatgc 186
 DB 1139 tcatcccaagg 1150

Db 992 aagacagagccacatgccctacacagatgctgtgtgacagaggtccagagatacatg 1051
 Oy 187 acctctcccccacagcctgcccatgacgtgacattaaattcagaactatc 246
 Db 1052 acctctcccccacagcctgcccatgacgtgacattaaattcagaactatc 1111
 Oy 247 tcatctccaagg 258
 Db 1112 tcatctccaagg 1123

RESULT 10

AA119925
 ID AA119925 standard; DNA; 1845 BP.
 AC AA119925;
 XX
 DT 11-JUN-1999 (first entry)
 XX
 DE Mammalian cytochrome P450 protein CYP2C9 encoding DNA.
 XX
 KW Bacterial; mammalian; cytochrome P450; chimeric; fusion protein;
 KW oxidase; hydrocarbon; carbon-hydrogen bond; hydroxylating;
 KW bioremediation; environmental pollutant; ss.
 XX
 OS Mammalia.
 XX
 PM MO9908812-A1.
 XX
 PD 25-FEB-1999.
 XX
 PF 17-AUG-1998; 98MO-US16979.
 XX
 PR 20-AUG-1997; 970S-0056754.
 XX
 PA (UTRP) UNIV ROCHESTER.
 XX
 PI Jones JP, Shimoji M;
 XX
 DR MPI: 1999-190131/16.
 DR P-PSDB: AAY04127.
 XX

XX New P450 fusion proteins - comprising a portion of a bacterial
 PT cytochrome P450 protein and a portion of a mammalian cytochrome P450
 PT protein.

XX Disclosure; Page 10-11: 51pp; English.

XX The present invention describes a fusion proteins comprising a portion
 CC of a bacterial cytochrome P450 protein and also a portion of a mammalian
 CC cytochrome P450 protein. The fusion protein can oxidise hydrocarbons or
 CC any compound having a carbon-hydrogen bond. The fusion protein can be
 CC used for hydroxylating a compound to be oxidised. It can also be used in
 CC the bioremediation of an environmental pollutant. Since the fusion
 CC protein is soluble, it can be subject to structural elucidation by X-ray
 CC crystallography for designing functional proteins. It can be readily
 CC expressed in soil bacteria to facilitate bioremediation. The present
 CC sequence encodes mammalian cytochrome P450 protein CYP2C9 from the
 CC present invention.

XX Sequence 1845 BP; 507 A; 424 C; 380 G; 534 T; 0 other;

Query Match 58.9%; Score 190.4; DB 20; Length 1845;
 Best Local Similarity 99.5%; Pred. No. 2.4e-51;
 Matches 191; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 67 cagctaaatccaggaagatgacgtgtgattggcagaacccgagccctgcattc 136
 Db 969 cagctaaatccaggaagatgacgtgtgattggcagaacccgagccctgcattc 1028
 Oy 127 aagcagagccacatgccctacacagatgctgtgtgacagaggtccagagatacatg 186

Db 1029 aagcagagccacatgccctacacagatgctgtgtgacagaggtccagagatacatg 1088
 Oy 187 acctctcccccacagcctgcccatgacgtgacattaaattcagaactatc 246
 Db 1089 acctctcccccacagcctgcccatgacgtgacattaaattcagaactatc 1148
 Oy 247 tcatctccaagg 258
 Db 1149 tcatctccaagg 1160

RESULT 11

AA111381
 ID AA111381 standard; cDNA; 1852 BP.
 AC AA111381;
 XX
 DT 09-SEP-1996 (first entry)
 XX
 DE Cytochrome P450 2C9 clone 65 coding sequence.
 XX
 KW Cytochrome P450 2C19; human; liver; PCR; primer; detection; CYP2C19;
 KW stereospecific S-mephenytoin 4'-hydroxylase activity; polymorphism; ss.
 XX
 OS Homo sapiens.

XX Key Location/Qualifiers
 FF CDS 11..1483
 FT /tag- a
 FT /product- Cytochrome P450 C9 clone 25
 FT misc_difference 1483
 FT /tag- b
 FT /note- *Variable position*
 FT misc_difference 1435
 FT /tag- c
 FT /note- *Variable position*

XX MO9530766-A1.

XX 16-NOV-1995.

XX 08-MAY-1995; 95MO-US05744.

XX 06-MAY-1994; 94US-0238821.

XX (USSH) US DEPT HEALTH & HUMAN SERVICES.

XX De Morais SMF, Goldstein JA, Romkes-Sparks M;
 XX MPI: 1996-077357/08.

XX P-PSDB: AAR9865.

XX New isolated cytochrome P450 2C subfamily member - used for
 PT identifying drugs metabolised by S-mephenytoin 4'-hydroxylase
 PT activity and to develop other screening assays

XX Example 2; Page 104; 169pp; English.

XX The sequences given in AA111378-81 encode allelic variants of cytochrome
 CC P450 2C9. The majority of clones isolated from liver S33 coded for 2C9.
 CC Of the clones encoding 2C9, only two allelic variants were found. 39
 CC of the 2C9 clones were identical with clone 67, and 11 were identical
 CC with clone 25. Clones 25 and 65 encode 25 and 65 amino acid and
 CC 3'-noncoding regions but contained 2 single base changes at positions
 CC 1075 and 1425. One of these base changes was conservative but the other
 CC results in one amino acid difference at position 359, Ile to Leu. The
 CC clones 29c and 6b differ by one nucleotide in the coding region.
 CC at position 1154, which results in a single amino acid change, Thr to Met
 CC and a polyadenylation signal 21 bases from the poly-A tail. Clone 6b
 CC has an unusually long 3'-noncoding region containing three possible
 CC polyadenylation signals with no poly-A tail. The differences in the 3'
 CC non coding regions could represent alternate splicing, allelic

CC variants, or possibly separate genes. These clones are designated as
 CC allium, and 2C18 because they differ by only one base in the
 CC coding region. They are most similar to 2C9 (82% amino acid homology)
 CC and 2C19 (81% amino acid homology).

XX Sequence 1852 BP: 514 A: 424 C: 380 G: 534 T: 0 other:

Query Match 58.9% Score 190.4; DB 17; Length 1852;
 Best Local Similarity 99.5%; Pred. No. 2.4e-51;
 Matches 191; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 OY 67 cagctaaagtcacaggaagattgacgtgtgattgacgaacacgagccctgcatgc 126
 DB 969 cagctaaagtcacaggaagattgacgtgtgattgacgaacacgagccctgcatgc 1028
 OY 127 aagacagggccatgcctcacacagatgctgtgtgacagaggtccagagatacttg 186
 DB 1029 aagacagggccatgcctcacacagatgctgtgtgacagaggtccagagatacttg 1088
 OY 187 accttctccacacagctcccatgcagtcagtcagtcagtcagtcagtcagtcagtc 246
 DB 1089 accttctccacacagctcccatgcagtcagtcagtcagtcagtcagtcagtcagtc 1148
 OY 247 tcattcccaagg 258
 DB 1149 tcattcccaagg 1160

RESULT 12

AAH51162 standard; cDNA; 1852 BP.

XX AAH51162:
 XX AC AAH51162:
 XX DT 29-AUG-2001 (first entry)
 XX DE Human CYP2C8 related DNA containing a biallelic polymorphism SEQ ID 53.
 XX KW Human: biallelic marker: single nucleotide polymorphism: SNP: MGST11;
 XX KW cytochrome P450; drug screening; S-mephenytoin 4'-hydroxylase;
 XX KW identification: mutagenic; carcinogenic; cytotoxic; haemoprotein;
 XX KW xenobiotic; environmental pollutant; ss.
 XX KW Homo sapiens.
 XX OS US5786191-A.
 XX PF 28-JUL-1998.
 XX PD 22-FEB-1994; 94US-0201118.
 XX PR 22-FEB-1994; 94US-0201118.
 XX PR 09-APR-1992; 92US-0864962.
 XX PA (GOLD)/ GOLDSTEIN J A.
 XX PA (ROM/)/ ROMKES-SPARKS M.
 XX PT Goldstein JA, Romkes-sparks M;
 XX DR WPI: 1998-436528/37.
 XX PT Screening for drugs metabolised by cytochrome P450 - for identifying
 XX PT mutagenic, carcinogenic, or cytotoxic compounds

XX Example 2: Column 53-56; 63pp: English.

XX This sequence encodes a human cytochrome P450 2C9 polypeptide isolated
 CC from clone 65. This polypeptide is a member of the cytochrome P450
 CC family, which is a group of proteins that are involved in the metabolism of
 CC 4'-hydroxylase activity. The protein can also be used to identify a
 CC mutagenic, carcinogenic or cytotoxic compound. Cytochrome P450 are a
 CC large family of haemoprotein enzymes capable of metabolising xenobiotics
 CC such as drugs carcinogens and environmental pollutants as well as

CC endobiotics such as steroids, fatty acids and prostaglandins.

XX Sequence 1852 BP: 514 A: 424 C: 380 G: 534 T: 0 other:

Query Match 58.9% Score 190.4; DB 19; Length 1852;
 Best Local Similarity 99.5%; Pred. No. 2.4e-51;
 Matches 191; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 OY 67 cagctaaagtcacaggaagattgacgtgtgattgacgaacacgagccctgcatgc 126
 DB 969 cagctaaagtcacaggaagattgacgtgtgattgacgaacacgagccctgcatgc 1028
 OY 127 aagacagggccatgcctcacacagatgctgtgtgacagaggtccagagatacttg 186
 DB 1029 aagacagggccatgcctcacacagatgctgtgtgacagaggtccagagatacttg 1088
 OY 187 accttctccacacagctcccatgcagtcagtcagtcagtcagtcagtcagtcagtc 246
 DB 1089 accttctccacacagctcccatgcagtcagtcagtcagtcagtcagtcagtcagtc 1148
 OY 247 tcattcccaagg 258
 DB 1149 tcattcccaagg 1160

RESULT 13

AAH51162 standard; DNA; 1001 BP.

XX AAH51162:
 XX AC AAH51162:
 XX DT 29-AUG-2001 (first entry)
 XX DE Human CYP2C8 related DNA containing a biallelic polymorphism SEQ ID 53.
 XX KW Human: biallelic marker: single nucleotide polymorphism: SNP: MGST11;
 XX KW cytochrome P450; drug screening; S-mephenytoin 4'-hydroxylase;
 XX KW identification: mutagenic; carcinogenic; cytotoxic; haemoprotein;
 XX KW xenobiotic; environmental pollutant; ss.
 XX KW Homo sapiens.
 XX OS WO200058508-A2.
 XX PF 05-OCT-2000.
 XX PD 24-MAR-2000; 2000WO-IB00403.
 XX PR 25-MAR-1999; 99US-0126269.
 XX PR 30-APR-1999; 99US-0131961.
 XX PA (GEST) GENSET.
 XX PT Blumenfeld M, Bougueleret L, Chumakov I, Cohen-Akenine A;
 XX DR WPI: 2000-638353/61.
 XX PT Polynucleotides comprising sequences from malate decarboxylase
 XX PT enzyme-related biallelic markers used for genotyping -

XX Claim 13: Page 283; 673pp: English.

XX Sequences AAH51110-AAH51593 represent human DNA fragments which contain
 CC biallelic markers. The sequences are related to various human genes
 CC including microsomal glutathione S-transferase II (MGSTII), malate
 CC decarboxylase enzyme (DMEI/MEI), cytochrome P450, glutathione
 CC reductase/synthase (GSHR/GSHS), flavin-containing monooxygenases (FMO),
 CC gamma-glutamyltransferase 5 (GGT5), dipeptidase (DP), glucose 6-phosphate

dehydrogenase (G6PDH), phosphoglucose dehydrogenase (G6PDH), and uridine diphosphate-glucosyl transferase (UGT2). Each of these sequences contains a 11 base pair region which is represented in the sequence as a degenerate/undefined region. The region which the allelic marker containing sequences are related are involved in drug metabolism. Sequences AAH51594 - AAH51598 represent the genomic sequence of the MESTII gene and four alternative MESTII cDNA sequences. AAH52905-AAH52906 are MESTII gene products. PCR primers AAH51599 and AAH51600 are used in an example for the amplification of human genomic DNA fragments. The invention includes a method of genotyping comprising determining the identity of a nucleotide at a MESTII-related allelic marker in a population of an individual for inclusion in clinical trials for a drug treatment. The method is also used to detect association between a drug allele and phenotype, and to detect association between haplotype and phenotype. The polynucleotides are used, in hybridization assays, sequencing assays or allele specific amplification assays. The method can be used to determine whether an individual suffers or is at risk of developing asthma or is at risk of developing hepatotoxicity on treatment with zileuton.

Sequence 1001 BP; 251 A; 180 C; 193 G; 366 T; 11 other;

Query Match 58.5%; Score 189; DB 21; Length 1001;
Best Local Similarity 93.3%; Pred. No. 5, 1e-51;
Matches 251; Conservative 0; Mismatches 9; Indels 9; Gaps 5;
OY 64 tatcagctaaagtcacagagagatgacgtgtgattgacagaaacccgagccctgca 123
DB 1 tatcagctaaagtcacagagagatgacgtgtgattgacagaaacccgagccctgca 60
OY 124 tccacagacagagacacacccctacacagatgctgtgtgacagagatccagatcc 183
DB 61 tgcacagacagagacacacccctacacagatgctgtgtgacagagatccagatcc 120
OY 184 ttgacctcttccacacagctgcccctacagatgacctgtgacattaaatcagaaa-- 241
DB 121 ttgacctcttccacacagctgcccctacagatgacctgtgacattaaatcagaaa 11
OY 242 -ctatctcttccacacagctgcccctacagatgacctgtgacattaaatcagaaaac 180
DB 181 antatctcttccacacagctgcccctacagatgacctgtgacattaaatcagaaaac 297
OY 298 a--gtcccaatctaatatcatctttt 323
DB 241 aagtcacacaaatctaatatcatctttt 269

RESULT 14

AAH22716
ID AAX22716 standard; DNA: 1419 BP.
AC AAX22716;
XX 27-MAY-1999 (first entry)
XX Human cytochrome P450 2C10 variant DNA.
DE
XX Cytochrome P450; detergent; enzyme inhibitor; non-ionic; ionic;
KW purification; drug oxidation; steroid; carcinogen; pesticide; human; ss.
OS Homo sapiens.
XX US:936157-A.
XX 23-MAR-1999.
XX 10-FEB-1994; 94US-0194981.
XX 10-FEB-1994; 94US-0194981.
XX

PA (UTVA-) UNIV VANDERBILT.
XX Gilliam EMJ, Guengerich FP, Guo Z, Sandhu P;
XX WPI; 1999-228609/19.
XX
XX Purifying recombinant cytochrome P450 - by utilising novel
PT combinations of detergents and enzyme inhibitors
XX
XX Disclosure: Column 55-56; 91pp; English.

CC This invention describes a recombinant cytochrome P450 protein which
CC is purified from a host cell culture using a combination of detergents
CC and enzyme inhibitors. The method comprises (a) fractionating the host
CC cells to prepare their membranes, (b) adding a non-ionic detergent to
CC detergents in a concentration of 0.8% to 2% (w/v) in a
CC detergent to the membranes between 2:1 to 10:1, (c) adding an ionic
CC detergent to the membranes between 4:1 to 4:1 (d) adding the
CC membrane-detergent mixture to remove insoluble materials and (e)
CC purifying the protein through a diethylaminoethyl-beaded column, then
CC through a carboxymethyl-beaded column, and finally through a
CC hydroxylapatite column. The method is used to purify cytochrome P450
CC proteins which are responsible for catalysing the oxidation of drugs,
CC steroids, carcinogens and pesticides. The method simplifies the
CC purification of P450 proteins, by the use of improved expression vectors
XX and novel detergent combinations.

Sequence 1419 BP; 409 A; 332 C; 307 G; 371 T; 0 other;

Query Match 58.5%; Score 188.8; DB 20; Length 1419;
Best Local Similarity 99.0%; Pred. No. 6, 9e-51;
Matches 190; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
OY 67 cagctaaagtcacagagagatgacgtgtgattgacagaaacccgagccctgcatgc 126
DB 905 cagctaaagtcacagagagatgacgtgtgattgacagaaacccgagccctgcatgc 964
OY 127 aagacagagacacacacccctacacagatgctgtgtgacagagatccagatccctg 186
DB 965 aagacagagacacacacccctacacagatgctgtgtgacagagatccagatccctg 1024
OY 187 acctcttccacacagctgcccctacagatgacctgtgacattaaatcagaaactatc 246
DB 1025 acctcttccacacagctgcccctacagatgacctgtgacattaaatcagaaactatc 1084
OY 247 tcatctccagag 258
DB 1085 tcatctccagag 1096

RESULT 15

AAH22715
ID AAX22715 standard; DNA: 1591 BP.
AC AAX22715;
XX 27-MAY-1999 (first entry)
XX Human cytochrome P450 2C10 DNA.
DE
XX Cytochrome P450; detergent; enzyme inhibitor; non-ionic; ionic;
KW purification; drug oxidation; steroid; carcinogen; pesticide; human; ss.
OS Homo sapiens.
XX US5886157-A.
XX 23-MAR-1999.
XX 10-FEB-1994; 94US-0194981.
XX

PR 10-FEB-1994; 94US-0194981.
XX (UTVA-) UNIV VANDERBILT.
XX Gillam EMJ, Guengerich FP, Guo Z, Sandhu P;
XX WPI; 1999-228609/19.
XX
PT Purifying recombinant cytochrome P450 - by utilizing novel
PT combinations of detergents and enzyme inhibitors
XX
PS Disclosure; Column 53-54; 91pp; English.
XX
CC This invention describes a recombinant cytochrome P450 protein which
CC is purified from a host cell culture using a combination of detergents
CC and enzyme inhibitors. The method comprises (a) fractionating the host
CC cells to prepare their membranes, (b) adding a non-ionic detergent to
CC the membranes in a concentration of 0.8% to 2% (w/v) in a
CC detergent-to-membrane ratio of between 0.8% and 0.8% (w/v) in a
CC detergent-to-membrane ratio of between 2:1 to 4:1 (d) centrifuging the
CC membrane-detergent mixture to remove insoluble materials and (e)
CC purifying the protein through a diethylaminoethyl-beaded column, then
CC through a carboxymethyl-beaded column, and finally through a
CC hydroxylapatite column. The method is used to purify cytochrome P450
CC proteins which are responsible for catalysing the oxidation of drugs,
CC steroids, carcinogens and pesticides. The method simplifies the
CC purification of P450 proteins, by the use of improved expression vectors
CC and novel detergent combinations.
XX
XX Sequence 1591 BP; 443 A; 372 C; 353 G; 423 T; 0 other;

Query Match 58.5%; Score 188.8; DB 20; Length 1591;
Best Local Similarity 99.0%; Pred. No. 7.3e-51;
Matches 190; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 67 cagctaaagtcacagagattgaacgtgtgattggcagaacccgagccctgcatgc 126
DB 959 cagctaaagtcacagagattgaacgtgtgattggcagaacccgagccctgcatgc 1018
QY 127 aagacagagccacatgcctcacagatgcgtggtgcacgaggtccagatccttg 186
DB 1019 aagacagagccacatgcctcacagatgcgtggtgcacgaggtccagatccttg 1078
QY 187 acctctctccaccagctgcccatgcagtgacctgtgacattaaattccagaactatc 246
DB 1079 acctctctccaccagctgcccatgcagtgacctgtgacattaaattccagaactatc 1138
QY 247 tcaattcccaag 258
DB 1139 tcaattcccaag 1150

Search completed: April 19, 2002, 08:15:33
Job time: 3233 sec

TITLE Creation of Genome-wide Protein Expression Libraries using Random
JOURNAL Nat. Biotechnol. 19 (5), 440 (2001) In press
COMMENT Activation of Genome-wide Protein Expression Libraries using Random
CONTACT Scott J. Cain
ADDRESS Athersys, Inc.
3201 Carnegie Ave. Cleveland, OH 44115, USA
Phone: 216 361 9500
Fax: 216 361 9596
Email: scain@atersys.com
FEATURES High quality sequence stop: 433.
Location/Qualifiers
1. .901
/organism="Homo sapiens"

```

/obj_xref=taxon:9606
/clone_lib=AtHERsys RAGE Library*
/cell_line=H1080*
/note=See 'Creation of Genome-wide Protein Expression
Libraries using Random Activation of Gene Expression', the
Nature Biotechnology, in press. Note that even though, the

```

```

BASE COUNT      230 a      241 c      199 g      231 t
ORIGIN
Query Match      58.9%      Score 190.4      DB 11:      Length 901:
Best Local Similarity 99.5%      Pred. No. 5, 2e-41
Matches 191:      Conservative 0:      Mismatches 41:      Indels 0:      Gaps
67  cagctaaagctccaggaagaattacagctgattgcaggaacccgagccctcgcgc 126
148  CAGCTAAAGCTCCAGGAAGATTACAGCTGATTGTCGAGAACCCGAGCCCTCGATGC 207

```

Qy	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385	386	387	388	389	390	391	392	393	394	395	396	397	398	399	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420	421	422	423	424	425	426	427	428	429	430	431	432	433	434	435	436	437	438	439	440	441	442	443	444	445	446	447	448	449	450	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466	467	468	469	470	471	472	473	474	475	476	477	478	479	480	481	482	483	484	485	486	487	488	489	490	491	492	493	494	495	496	497	498	499	500	501	502	503	504	505	506	507	508	509	510	511	512	513	514	515	516	517	518	519	520	521	522	523	524	525	526	527	528	529	530	531	532	533	534	535	536	537	538	539	540	541	542	543	544	545	546	547	548	549	550	551	552	553	554	555	556	557	558	559	560	561	562	563	564	565	566	567	568	569	570	571	572	573	574	575	576	577	578	579
----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

SOURCE	human sapiens
ORGANISM	Eukaryota; Metazoa; Chordata; Cranial; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE	(bases 1 to 3076)
AUTHORS	Harrington, J.L., Thornton, B., Rundlett, S., Jackson, P.D., Perry, R., Wilson, D.J., Shorrock, M., Ramachandran, R., Whittington, J., Veloso, J., Krashinsky, K., Clark, S., Wasylyuk, R., Smith, E., Calos, N., Hess, J., Cotman, K., Lo, K., Offenbacher, J., Danzig, J., and Dincer, M.
TITLE	Creation of Genome-wide Protein Expression Libraries using Random
JOURNAL	Activation of Gene Expression
COMMENT	Nat. Biotechnol. 19 (5): 440 (2001) In press Contact: Scott J. Cain Athersys, Inc. 3201 Carnegie Ave, Cleveland, OH 44115, USA Tel: 216 431 9900 Fax: 216 361 9596 Email: scjain@atersys.com

FEATURES High quality sequence stop: 354.

source

Location/Qualifiers
1. 907
/db_xref=taxon:9606
/organism="Homo sapiens"
/cell_line="HT1080"
/note="See 'Creation of Genome-wide Protein Expression Libraries using Random Activation of Gene Expression', Nature Biotechnology, in press. Note that even though the cell type indicated is HT1080, since a random activation method was used, these sequence tags are not necessarily expressed in HT1080 under normal circumstances."

BASE COUNT 226 a 234 c 192 g 254 t 1 others
ORIGIN

Query Match 58.9% Score 190.4; DB 11; Length 907;
Best Local Similarity 99.5%; Pred. No. 5.2e-41;
Matches 191; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 67 cagctaaagtcaggagagattgaacgtggtggtgacgaacccgagccctgcacgc 126
|||||
Db 145 CAGCTAAAGTCAGGAGAGATTGACGTGTGTGGCAGAAACCGAGCCCTGCATGC 204
Oy 127 agacagagccacatgcccctacacagatgctggtgacagaggtccagagataccttg 186
|||||
Db 205 AGACAGAGCCACATGCCCCTACACAGATGCTGTGTGGCAGAGTCCAGAGATACATG 264
Oy 187 acccttcctccacacagctccctccctcagctgacattgaattcagaactatc 246
|||||
Db 265 ACCTTCTCCCCACAGAGCTCCCTCCCATGCACTGTGACATTAAATTTCAGAACTATC 324
Oy 247 tcattcccaagg 258
|||||
Db 325 TCATTCCCAAGG 336

RESULT 9

LOCUS

DEFINITION BGI98889 909 bp mRNA EST 21-APR-2001
ACCESSION RST18160 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
VERSION BGI98889
KEYWORDS EST.

ORIGIN human
human

REFERENCE 1
AUTHORS Harrington, J.J., Sherf, B., Rundlett, S., Jackson, P.D., Perry, R., Cain, S., Dahl, T., Thornton, M., Ramachandran, R., Whittington, J., Lerner, L., Krashoc, D., McElligott, K., Clark, S., Mays, R., Smith, E., Veloso, M., Hess, J., Cothren, K., Lo, K., Offenbacher, J., Danzig, J., and Ducar, M.
TITLE Creation of Genome-wide Protein Expression Libraries using Random Activation of Gene Expression
JOURNAL Nat. Biotechnol. 19 (5), 440 (2001) In press
COMMENT Contact: Scott J. Cain
3201 Carnegie Ave., Cleveland, OH 44115, USA
Tel: 216 431 9900
Fax: 216 361 9596
Email: scaif@atersys.com
High quality sequence stop: 532.

FEATURES

source

Location/Qualifiers
1. 909
/db_xref=taxon:9606
/organism="Homo sapiens"
/cell_line="HT1080"
/note="See 'Creation of Genome-wide Protein Expression Libraries using Random Activation of Gene Expression', Nature Biotechnology, in press. Note that even though the cell type indicated is HT1080, since a random activation method was used, these sequence tags are not necessarily expressed in HT1080 under normal circumstances."

FEATURES High quality sequence stop: 532.

source

Location/Qualifiers
1. 909
/db_xref=taxon:9606
/organism="Homo sapiens"
/cell_line="HT1080"
/note="See 'Creation of Genome-wide Protein Expression Libraries using Random Activation of Gene Expression', Nature Biotechnology, in press. Note that even though the cell type indicated is HT1080, since a random activation method was used, these sequence tags are not necessarily expressed in HT1080 under normal circumstances."

BASE COUNT 237 a 240 c 181 g 250 t 1 others
ORIGIN

Query Match 58.9% Score 190.4; DB 11; Length 909;
Best Local Similarity 99.5%; Pred. No. 5.2e-41;
Matches 191; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 67 cagctaaagtcaggagagattgaacgtggtggtgacgaacccgagccctgcacgc 126
|||||
Db 147 CAGCTAAAGTCAGGAGAGATTGACGTGTGTGGCAGAAACCGAGCCCTGCATGC 206
Oy 127 agacagagccacatgcccctacacagatgctggtgacagaggtccagagataccttg 186
|||||
Db 207 AGACAGAGCCACATGCCCCTACACAGATGCTGTGTGGCAGAGTCCAGAGATACATG 266
Oy 187 acccttcctccacacagctccctccctcagctgacattgaattcagaactatc 246
|||||
Db 267 ACCTTCTCCCCACAGAGCTCCCTCCCATGCACTGTGACATTAAATTTCAGAACTATC 326
Oy 247 tcattcccaagg 258
|||||
Db 327 TCATTCCCAAGG 338

RESULT 10

LOCUS

DEFINITION BGI84407 388 bp mRNA EST 21-APR-2001
ACCESSION RST3332 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
VERSION BGI84407
KEYWORDS EST.

ORIGIN human
human

REFERENCE 1
AUTHORS Harrington, J.J., Sherf, B., Rundlett, S., Jackson, P.D., Perry, R., Cain, S., Dahl, T., Thornton, M., Ramachandran, R., Whittington, J., Lerner, L., Krashoc, D., McElligott, K., Clark, S., Mays, R., Smith, E., Veloso, M., Hess, J., Cothren, K., Lo, K., Offenbacher, J., Danzig, J., and Ducar, M.
TITLE Creation of Genome-wide Protein Expression Libraries using Random Activation of Gene Expression
JOURNAL Nat. Biotechnol. 19 (5), 440 (2001) In press
COMMENT Contact: Scott J. Cain
3201 Carnegie Ave., Cleveland, OH 44115, USA
Tel: 216 431 9900
Fax: 216 361 9596
Email: scaif@atersys.com
High quality sequence stop: 338.

FEATURES

source

Location/Qualifiers
1. 388
/organism="Homo sapiens"
/db_xref=taxon:9606
/cell_line="HT1080"
/note="See 'Creation of Genome-wide Protein Expression Libraries using Random Activation of Gene Expression', Nature Biotechnology, in press. Note that even though the cell type indicated is HT1080, since a random activation method was used, these sequence tags are not necessarily expressed in HT1080 under normal circumstances."

FEATURES High quality sequence stop: 338.

source

Location/Qualifiers
1. 388
/organism="Homo sapiens"
/db_xref=taxon:9606
/cell_line="HT1080"
/note="See 'Creation of Genome-wide Protein Expression Libraries using Random Activation of Gene Expression', Nature Biotechnology, in press. Note that even though the cell type indicated is HT1080, since a random activation method was used, these sequence tags are not necessarily expressed in HT1080 under normal circumstances."

BASE COUNT

ORIGIN

Location/Qualifiers
1. 388
/db_xref=taxon:9606
/organism="Homo sapiens"
/cell_line="HT1080"
/note="See 'Creation of Genome-wide Protein Expression Libraries using Random Activation of Gene Expression', Nature Biotechnology, in press. Note that even though the cell type indicated is HT1080, since a random activation method was used, these sequence tags are not necessarily expressed in HT1080 under normal circumstances."

Query Match

58.5% Score 188.8; DB 11; Length 388;

```
Best Local Similarity 99.0%; Pred. No. 1.3e-40; Mismatches 0; Indels 2; Gaps 0;
Matches 190; Conservative 0;

Oy 67 cagctaaatccagagagattgagctggtgattgagcagaacggagccctgcacgc 126
Db 182 CAGCTAAAGTCCAGAGAGATTGAGCTGTGTGATGGCAGAAACCGAGCCCTGCATGC 241

Oy 127 aagcagagagcagatgcctcacaagatgctggtgacagaggtccagagacacctg 186
Db 242 AAGCAGAGAGCAGATGCCCTCACAGATGCTGTGTGACAGAGGTCCAGAGATCATTTG 301

Oy 187 acctctccacacagctccctcagcagctgacattgacattaaattcagaactatc 246
Db 302 ACCTCTCCACACAGCTCCCTCAGCAGCTGACCTGTGACATTAATTTCAGAACTATC 361

Oy 247 tcattcccaagg 258
Db 362 TCATTCCCAAGG 373

RESULT 11
LOCUS AV651261 620 bp mRNA EST 07-SEP-2000
DEFINITION AV651261 GLC Homo sapiens cDNA clone GLC1261 3', mRNA sequence.
ACCESSION AV651261
VERSION AV651261.1 GI:9872275
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 620)
AUTHORS Qian,B., Wu,T., Huang,O., Huang,C., Kang,B., Gao,X., Xu,Z., Xiao,H.,
, Xu,X., Li,N., Peng,Y., Liu,F., Ou,J., Song,H., Cheng,Z., Ou,J.,
Zeng,L., Xu,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu
, G., Yang,Y., Gu,Y., Chen,Z. and Han,Z.
TITLE Homo sapiens cDNA clone
JOURNAL Unpublished (2000)
COMMENT Contact: Zengqiang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801913(ex.45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.
FEATURES
source
1..620
/organism="Homo sapiens"
/clone="GLC1261"
/clone_lib="GLC"
/tissue_type="corresponding non cancerous liver tissue"
/dev_stage="Adult"
/lab_host="SOLR"
/Note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
XhoI"
BASE COUNT 205 a 141 c 112 g 162 t
ORIGIN
Query Match 58.5%; Score 188.8; DB 10; Length 620;
Best Local Similarity 99.0%; Pred. No. 1.3e-40;
Matches 190; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Oy 67 cagctaaatccagagagattgagctggtgattgagcagaacggagccctgcacgc 126
Db 379 CAGCTAAAGTCCAGAGAGATTGAGCTGTGTGATGGCAGAAACCGAGCCCTGCATGC 438
Oy 127 aagcagagagcagatgcctcacaagatgctggtgacagaggtccagagacacctg 186
Db 439 AAGCAGAGAGCAGATGCCCTCACAGATGCTGTGTGACAGAGGTCCAGAGATCATTTG 498

Best Local Similarity 99.0%; Pred. No. 1.3e-40; Mismatches 0; Indels 2; Gaps 0;
Matches 190; Conservative 0;

Oy 187 acctctccacacagctccctcagcagctgacattgacattaaattcagaactatc 246
Db 499 ACCTCTCCACACAGCTCCCTCAGCAGCTGACCTGTGACATTAATTTCAGAACTATC 558

Oy 247 tcattcccaagg 258
Db 559 TCATTCCCAAGG 570

RESULT 12
LOCUS AV652420 661 bp mRNA EST 07-SEP-2000
DEFINITION AV652420 GLC Homo sapiens cDNA clone GLC2420 3', mRNA sequence.
ACCESSION AV652420
VERSION AV652420.1 GI:9873434
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 661)
AUTHORS Qian,B., Wu,T., Huang,O., Huang,C., Kang,B., Gao,X., Xu,Z., Xiao,H.,
, Xu,X., Li,N., Peng,Y., Liu,F., Ou,J., Song,H., Cheng,Z., Ou,J.,
Zeng,L., Xu,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu
, G., Yang,Y., Gu,Y., Chen,Z. and Han,Z.
TITLE Homo sapiens cDNA clone
JOURNAL Unpublished (2000)
COMMENT Contact: Zengqiang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801913(ex.45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.
FEATURES
source
1..661
/organism="Homo sapiens"
/clone="GLC2420"
/clone_lib="GLC"
/tissue_type="corresponding non cancerous liver tissue"
/dev_stage="Adult"
/lab_host="SOLR"
/Note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
XhoI"
BASE COUNT 166 a 189 c 130 g 175 t
ORIGIN
Query Match 58.5%; Score 188.8; DB 10; Length 661;
Best Local Similarity 99.0%; Pred. No. 1.4e-40;
Matches 190; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Oy 67 cagctaaatccagagagattgagctggtgattgagcagaacggagccctgcacgc 126
Db 52 CAGCTAAAGTCCAGAGAGATTGAGCTGTGTGATGGCAGAAACCGAGCCCTGCATGC 111
Oy 127 aagcagagagcagatgcctcacaagatgctggtgacagaggtccagagacacctg 186
Db 112 AAGCAGAGAGCAGATGCCCTCACAGATGCTGTGTGACAGAGGTCCAGAGATCATTTG 171
Oy 187 acctctccacacagctccctcagcagctgacattgacattaaattcagaactatc 246
Db 172 ACCTCTCCACACAGCTCCCTCAGCAGCTGACCTGTGACATTAATTTCAGAACTATC 231
Oy 247 tcattcccaagg 258
Db 232 TCATTCCCAAGG 243

RESULT 13
BG217233
```

LOCUS BG1217233 802 bp mRNA EST 21-APR-2001
 DEFINITION Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
 ACCESSION BG1217233
 VERSION BG1217233.1 GI:13743754
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 REFERENCE 1 (bases 1 to 802)
 AUTHORS Harrington, J.J., Sherf, B., Rundlett, S., Jackson, P.D., Perry, R.,
 Cain, S., Dahl, T., Thornton, M., Ramachandran, R., Whittington, J.,
 Lerner, L., Krashinsky, D., McElligott, K., Clark, S., Mays, R., Smith, E.,
 Veloso, N., Hess, J., Cothren, K., Lo, K., Offenbacher, J., Danzig, J.,
 and Ducar, M.
 TITLE Creation of Genome-wide Protein Expression Libraries using Random
 Activation of Gene Expression
 JOURNAL Nat. Biotechnol. 19 (5), 440 (2001) In press
 COMMENT Contact: Scott J. Cain
 Athersys, Inc.
 3201 Carnegie Ave, Cleveland, OH 44115, USA
 Tel: 216 431 9900
 Fax: 216 361 9596
 Email: scain@atersys.com
 High quality sequence stop: 473.
 Location/Qualifiers
 1..802
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="Athersys RAGE Library"
 /cell_line="HT1080"
 /note="See 'Creation of Genome-wide Protein Expression
 Libraries using Random Activation of Gene Expression',
 Nature Biotechnology, in press. Note that even though the
 cell type indicated is HT1080, since a random activation
 method was used, these sequence tags are not necessarily
 expressed in HT1080 under normal circumstances."
 BASE COUNT 200 a 217 c 171 g 214 t
 ORIGIN
 Query Match 58.5% Score 188.8; DB 11; Length 802;
 Best Local Similarity 99.0%; Pred. No. 1.4e-40;
 Matches 190; Conservative 0; Mismatches 2; Indels 0; Gaps 0:
 QY 67 cagctaaagtcagagagattgaacgtgtgattgacgaacacgagccctgctgc 126
 Db 145 CAGCTAAAGTCCAGAGAGATTGAACGTGTGATTGACGAACACGAGCCCTGCATGC 204
 QY 127 aagcagagagacacatgccctacacagatgctgtgacgaggtccagagatacttg 186
 Db 205 AAGCAGAGAGACACATGCCCTACACAGATGCTGTGTGCGGAGGTCCAGAGATACATG 264
 QY 187 accttctcccaacagctgcccctacacagatgctgtgacgaggtccagagatacttc 246
 Db 265 ACCTTCTCCCAACAGCTGCCCTACACAGATGCTGTGTGCGGAGGTCCAGAGATACATG 324
 QY 247 taattcccaagg 258
 Db 325 TCATTCCTCCAGG 336
 RESULT 14
 LOCUS BG196822 876 bp mRNA EST 21-APR-2001
 DEFINITION Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
 ACCESSION BG196822
 VERSION BG196822.1 GI:13718509
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE 1 (bases 1 to 876)
 AUTHORS Harrington, J.J., Sherf, B., Rundlett, S., Jackson, P.D., Perry, R.,
 Cain, S., Dahl, T., Thornton, M., Ramachandran, R., Whittington, J.,
 Lerner, L., Krashinsky, D., McElligott, K., Clark, S., Mays, R., Smith, E.,
 Veloso, N., Hess, J., Cothren, K., Lo, K., Offenbacher, J., Danzig, J.,
 and Ducar, M.
 TITLE Creation of Genome-wide Protein Expression Libraries using Random
 Activation of Gene Expression
 JOURNAL Nat. Biotechnol. 19 (5), 440 (2001) In press
 COMMENT Contact: Scott J. Cain
 Athersys, Inc.
 3201 Carnegie Ave, Cleveland, OH 44115, USA
 Tel: 216 431 9900
 Fax: 216 361 9596
 Email: scain@atersys.com
 High quality sequence stop: 456.
 Location/Qualifiers
 1..876
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="Athersys RAGE Library"
 /cell_line="HT1080"
 /note="See 'Creation of Genome-wide Protein Expression
 Libraries using Random Activation of Gene Expression',
 Nature Biotechnology, in press. Note that even though the
 cell type indicated is HT1080, since a random activation
 method was used, these sequence tags are not necessarily
 expressed in HT1080 under normal circumstances."
 BASE COUNT 222 a 222 c 194 g 237 t 1 others
 ORIGIN
 Query Match 58.5% Score 188.8; DB 11; Length 876;
 Best Local Similarity 99.0%; Pred. No. 1.4e-40;
 Matches 190; Conservative 0; Mismatches 2; Indels 0; Gaps 0:
 QY 67 cagctaaagtcagagagattgaacgtgtgattgacgaacacgagccctgctgc 126
 Db 144 CAGCTAAAGTCCAGAGAGATTGAACGTGTGATTGACGAACACGAGCCCTGCATGC 203
 QY 127 aagcagagagacacatgccctacacagatgctgtgacgaggtccagagatacttg 186
 Db 204 AAGCAGAGAGACACATGCCCTACACAGATGCTGTGTGCGGAGGTCCAGAGATACATG 263
 QY 187 accttctcccaacagctgcccctacacagatgctgtgacgaggtccagagatacttc 246
 Db 264 ACCTTCTCCCAACAGCTGCCCTACACAGATGCTGTGTGCGGAGGTCCAGAGATACATG 323
 QY 247 taattcccaagg 258
 Db 324 TCATTCCTCCAGG 335
 RESULT 15
 LOCUS BG197837 896 bp mRNA EST 21-APR-2001
 DEFINITION Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
 ACCESSION BG197837
 VERSION BG197837.1 GI:13719524
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 REFERENCE 1 (bases 1 to 896)
 AUTHORS Harrington, J.J., Sherf, B., Rundlett, S., Jackson, P.D., Perry, R.,
 Cain, S., Dahl, T., Thornton, M., Ramachandran, R., Whittington, J.,
 Lerner, L., Krashinsky, D., McElligott, K., Clark, S., Mays, R., Smith, E.,
 Veloso, N., Hess, J., Cothren, K., Lo, K., Offenbacher, J., Danzig, J.,
 and Ducar, M.
 TITLE Creation of Genome-wide Protein Expression Libraries using Random
 Activation of Gene Expression
 JOURNAL Nat. Biotechnol. 19 (5), 440 (2001) In press

COMMENT

Contact: Scott J. Cain
NCS, Inc.
3201 Ely Road
Ave., Cleveland, OH 44115, USA
Tel: 216 431 9900
Fax: 216 361 9596
Email: scain@ncs.com
High quality sequence stop: 321.
Location/Qualifiers

FEATURES

source
1. .896
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="HT1080"
/note="See 'Creation of Genome-wide Protein Expression
Libraries using Random Activation of Gene Expression',
Nature Biotechnology, in press. Note that even though the
cell type indicated is HT1080, since a random activation
method was used, these sequence tags are not necessarily
expressed in HT1080 under normal circumstances."

BASE COUNT 221 a 251 c 181 g 242 t
ORIGIN

Query Match 58.5%; Score 188.8; DB 11; Length 896;
Best Local Similarity 99.0%; Pred. No. 1.4e-40;
Matches 190; Conservative 0; Mismatches 2; Indels 0; Gaps 0:
QY 67 cagctaaagtcacgaagatgacgtgattggcagaacccgagccctcgatgc 126
DB 114 CAGCTAAAGTCCGAGAGAGATGACGTGATTGGCAGAACCCGAGCCCTCGCATGC 173
QY 127 aaacacagagacacatgccctacacagatgctgtgtatccagagatccagatcccttg 186
DB 174 AAGACAGGACCCACATGCCCTACACAGATGCTGTGTGTCACGAGGTCCAGAGATACATTG 233
QY 187 accttctccaccacgcctgcccatgcagtgacacctgtgacattaaattccagaactatc 246
DB 234 AACTTCTCCACACAGCCTGCCCATGCTGACCTGTGACATTAAATTCAGAAACTATC 293
QY 247 tcatcccaagg 258
DB 294 TATTCCTCAGG 305

Search completed: April 19, 2002, 08:07:52
Job time: 2772 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OK nucleic - nucleic search, using sv model

Run on: April 19, 2002, 11:06:57 : Search time 232.27 Seconds
(without alignments)
163.811 Million cell updates/sec

Title: US-09-763-292-3

Perfect score: 333

Sequence: 1 cccggaattctacacaa.....aaattcatatcatatttt 323

Scoring table: IDENTITY.MUC

Gapop 10.0, Gapext 1.0

Searched: 78474 seqs, 5898373 residues

Total number of hits satisfying chosen parameters: 156948

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: /cgn2_6/ptodata/2/pna/PCT_NEM_COMB.seq.*
2: /cgn2_6/ptodata/2/pna/US06_NEW_COMB.seq.*
3: /cgn2_6/ptodata/2/pna/US08_NEW_COMB.seq.*
4: /cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq.*
5: /cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq.*
6: /cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq.*
7: /cgn2_6/ptodata/2/pna/US60_NEW_COMB.seq.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	62.8	19.4	591	6	US-10-106-698-858
2	56.8	17.6	526	6	US-10-106-698-2440
3	44.6	13.8	1186	6	US-10-103-313-621
C 4	44.6	13.8	1186	6	US-10-103-299-7204
C 5	38.6	12.0	1327	6	US-10-103-313-621
C 6	37.5	11.5	1182	6	US-10-103-313-622
C 7	37	11.5	1182	6	US-10-103-299-7203
C 8	31.4	9.7	1988	6	US-10-106-698-892
C 9	30.5	9.5	254	5	US-09-975-254-18019
C 10	30.2	9.3	35959	6	US-10-105-299-10167
C 11	29.4	9.1	589	6	US-10-015-219-101
C 12	29.4	9.1	1103	6	US-10-105-299-2978
C 13	29.4	9.1	4687	6	US-10-105-299-14978
C 14	29.4	9.1	17694	1	PCT-US02-05912-8
C 15	29.4	9.1	14653	5	US-09-975-254-26054
C 16	29.4	9.1	14653	5	US-09-975-254-26054
C 17	29.4	9.1	14653	5	US-09-975-254-26054
C 18	28.8	8.9	231	5	US-09-975-254-23269
C 19	28.8	8.9	1022	6	US-10-105-299-1724
C 20	28.8	8.9	14036	6	US-10-105-299-9008
C 21	28.8	8.9	14036	6	US-10-105-299-10472
C 22	28.8	8.9	14036	6	US-10-105-299-10518
C 23	28.6	8.9	257	5	US-09-975-254-5489
C 24	28.2	8.7	31169	6	US-10-105-299-9006
C 25	28	8.7	535	6	US-10-097-105-1335
C 26	28	8.7	535	6	US-10-097-105-1335

Sequence 12476, A
Sequence 4, Appl
Sequence 2, Appl
Sequence 1, Appl
Sequence 1234, A
Sequence 1035, A
Sequence 3035, A
Sequence 920, A
Sequence 120, A
Sequence 2163, A
Sequence 14218, A
Sequence 14221, A
Sequence 14224, A
Sequence 282, A
Sequence 907, A
Sequence 149, A
Sequence 14155, A

ALIGNMENTS

RESULT 1
US-10-105-698-858
: Sequence 858, Application US/10106698
: GENERAL INFORMATION:
: APPLICANT: Ruben et al
: TITLE: Human Tumor and Colon Cancer Associated Polynucleotides and Polypept
: FILE REFERENCE: PA00591
: CURRENT APPLICATION NUMBER: US/10/106.698
: PRIOR FILING DATE: 2002-03-27
: PRIOR APPLICATION NUMBER: PCT/US00/26524
: PRIOR FILING DATE: 2000-09-28
: PRIOR APPLICATION NUMBER: US 60/157.137
: PRIOR FILING DATE: 1999-09-29
: PRIOR APPLICATION NUMBER: US 60/163.280
: PRIOR FILING DATE: 1999-11-03
: NUMBER OF SEQ ID NOS: 856
: SOURCE: Saccharin Var. 3.0
: SEQ ID NO 858
: LENGTH: 591
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: misc.feature
: LOCATION: (132)..(132)
: OTHER INFORMATION: equals a.t.g. or c
: NAME/KEY: misc.feature
: LOCATION: (542)..(542)
: OTHER INFORMATION: n equals a.t.g. or c
: NAME/KEY: misc.feature
: LOCATION: (550)..(550)
: OTHER INFORMATION: n equals a.t.g. or c
: NAME/KEY: misc.feature
: LOCATION: (579)..(579)
: OTHER INFORMATION: equals a.t.g. or c
: NAME/KEY: misc.feature
: LOCATION: (590)..(590)
: OTHER INFORMATION: n equals a.t.g. or c
US-10-106-698-858

Query Match 19.4% Score 62.8; DB 6; Length 591;
Best Local Similarity 58.3% Pred. No. 3.1e-10;
Matches 109; Conservative 0; Mismatches 78; Indels 0; Gaps 0;
Oy 75 gtccacagagatgcacgctgattgacacgaggtccacgagatccttgccttctc 134
Db 7 gtccacagagatgcacgctgattgacacgaggtccacgagatccttgccttctc 66
Oy 135 agccatgccttcacacagatgctgattgacacgaggtccacgagatccttgccttctc 194

Db 67 getacatgctccatcacaccactcgcggtgattcagtgagcgccttggggacatcgc 126
Oy 195 cccaccagctccagcagtgacattgaattacgaataatcagaataatctctatccc 254
Db 127 ccttggtggtgaccatcagatcagcctcctgacatcgaagtcacaggggttcctcct 185
Oy 255 aaggtaa 261
Db 187 aaggaa 193

RESULT 2
US-10-106-698-7440
: Sequence ID: 106698
: GENERAL INFORMATION:
: APPLICANT: Ruben et al.
: TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptides
: FILE REFERENCE: PA0051
: CURRENT APPLICATION NUMBER: US/10/106,698
: PRIOR FILING DATE: 2002-03-27
: PRIOR FILING DATE: 2000-09-28
: PRIOR FILING DATE: 2000-09-28
: PRIOR FILING DATE: 1998-09-29
: PRIOR FILING DATE: 1999-11-03
: NUMBER OF SEQ ID NOS: 8564
: SOFTWARE: Patent In Ver. 3.0
: SEQ ID NO 2440
: LENGTH: 526
: TYPE: DNA
: ORGANISM: Homo sapiens
: NAME/KEY: misc_feature
: LOCATION: (41)..(41)
: OTHER INFORMATION: n equals a.t.g. or c
: NAME/KEY: misc_feature
: LOCATION: (23)..(23)
: OTHER INFORMATION: n equals a.t.g. or c
: NAME/KEY: misc_feature
: LOCATION: (32)..(32)
: OTHER INFORMATION: n equals a.t.g. or c
: NAME/KEY: misc_feature
: LOCATION: (310)..(310)
: OTHER INFORMATION: n equals a.t.g. or c
: NAME/KEY: misc_feature
: LOCATION: (332)..(332)
: OTHER INFORMATION: n equals a.t.g. or c
: NAME/KEY: misc_feature
: LOCATION: (336)..(336)
: OTHER INFORMATION: n equals a.t.g. or c
: NAME/KEY: misc_feature
: LOCATION: (375)..(375)
: OTHER INFORMATION: n equals a.t.g. or c
: NAME/KEY: misc_feature
: LOCATION: (387)..(387)
: OTHER INFORMATION: n equals a.t.g. or c
: NAME/KEY: misc_feature
: LOCATION: (414)..(414)
: OTHER INFORMATION: n equals a.t.g. or c
: NAME/KEY: misc_feature
: LOCATION: (419)..(419)
: OTHER INFORMATION: n equals a.t.g. or c
: NAME/KEY: misc_feature
: LOCATION: (463)..(463)
: OTHER INFORMATION: n equals a.t.g. or c

: NAME/KEY: misc_feature
: LOCATION: (198)..(198)
: OTHER INFORMATION: n equals a.t.g. or c
: NAME/KEY: misc_feature
: LOCATION: (518)..(518)
: OTHER INFORMATION: n equals a.t.g. or c
: NAME/KEY: misc_feature
: LOCATION: (523)..(523)
: OTHER INFORMATION: n equals a.t.g. or c
US-10-106-698-2440

Query Match 17.6% Score 56.8; DB 6; Length 526;
Best Local Similarity 63.0% Pred. No. 2e-08;
Matches 85; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

Oy 124 tgaagacagagacacagccctacacagatcgtgtgacagaggtccagagatacc 183
Db 15 ttcatacgaagcacaanngccatacagagcagctctatgagctccagggattt 74
Oy 184 ttaactctccac 243
Db 75 cagactctccac 134
Oy 244 atctcattcccaagg 258
Db 135 acatcatcccaagg 149

RESULT 3
US-10-103-313-621
: Sequence ID: 103313
: GENERAL INFORMATION:
: APPLICANT: Rosen et al.
: TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
: FILE REFERENCE: P0207C1
: CURRENT APPLICATION NUMBER: US/10/103,313
: CURRENT FILING DATE: 2002-03-12
: NUMBER OF SEQ ID NOS: 653
: SOFTWARE: Patent In Ver. 2.0
: SEQ ID NO 621
: LENGTH: 1186
: TYPE: DNA
: ORGANISM: Homo sapiens
US-10-103-313-621

Query Match 13.8% Score 44.6; DB 6; Length 1186;
Best Local Similarity 53.1% Pred. No. 0.00014;
Matches 95; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

Oy 80 ggaagagattgaacgtgattgacgaacacccgagccctcgtcagtcagagagagacca 139
Db 5 gcaggagctggacccctgtgtgagtgaggagcccccacacacacacacacacac 64
Oy 140 catgctctac 199
Db 65 cctgctctccac 124
Oy 200 cagctcctccac 258
Db 125 gggagctcggcgcacccctcaccctcaccacacacacacacacacacacacac 183

RESULT 4
US-10-105-299-7204/c
: Sequence ID: 105299
: GENERAL INFORMATION:
: APPLICANT: Rosen, et al.
: TITLE OF INVENTION: Human Secreted Proteins
: FILE REFERENCE: PS950
: CURRENT APPLICATION NUMBER: US/10/105,299

;; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;; FILE REFERENCE: P1207C1
;; CURRENT APPLICATION NUMBER: US/10/103,313
;; PRIOR APPLICATION REMOVED - See File Wrapper or Palm
;; NUMBER OF SEQ ID NOS: 653
;; SOFTWARE: PatentIn Ver. 2.0
;; SEQ ID NO 622
;; LENGTH: 1182
;; TYPE: DNA
;; ORGANISM: Homo sapiens
US-10-103-313-622

Query Match 11.5% Score 37; DB 6; Length 1182;
Best Local Similarity 50.3%; Pred. No. 0.03;
Matches 91; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

Oy 80 ggaagagattgaacgtgattgacgaacacgagccctgcatgcaagacagagagcca 139
Db 1182 GCAGGAGCTGGACCTCTGTGGTAGGGTGGAGGCGCCGCCCAAGCCTGGACTATCGCGTGTG 1123
Oy 140 catgacctacacagatgctggtgacgaggtccagagatgaccttgccttctcccccac 199
Db 1122 CTGCTCCCTAGCCGACGAGGCTGCTGGAGATCCAGTGTCTTCAAGCGTGGTGGCCCT 1063
Oy 200 cagctgccccatgacgtgacgtgacattaaattcagaatactatcttcccaagg 258
Db 1062 GGGCGCTGGCGCGACCTCCACCTCGACACCCAGCCTGACAGCCACTGTCTGCCCAAG 1004

RESULT 5
US-10-105-299-7201/c
;; Sequence 7201, Application US/10105299
;; GENERAL INFORMATION: a) al
;; TITLE OF INVENTION: Human Secreted Proteins
;; FILE REFERENCE: PS950
;; CURRENT APPLICATION NUMBER: US/10/105,299
;; CURRENT FILING DATE: 2002-03-26
;; NUMBER OF SEQ ID NOS: 15197
;; PRIOR APPLICATION REMOVED - See File Wrapper or Palm
;; SOFTWARE: PatentIn Ver. 2.0
;; SEQ ID NO 7201
;; LENGTH: 13327
;; TYPE: DNA
;; ORGANISM: Homo sapiens
US-10-105-299-7201

Query Match 12.0% Score 38.6; DB 6; Length 13327;
Best Local Similarity 50.8%; Pred. No. 0.028;
Matches 92; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

Oy 80 ggaagagattgaacgtgattgacgaacacgagccctgcatgcaagacagagagcca 139
Db 9351 GCAGGAGCTGGACCTCTGTGGTAGGGTGGAGGCGCCGCCCAAGCCTGGACTATCGCGTGTG 9292
Oy 140 catgacctacacagatgctggtgacgaggtccagagatgaccttgccttctcccccac 199
Db 9291 CTGCTCCCTAGCCGACGAGTGTGCTGGAGATCCAGTGTCTTCAAGCGTGGTGGCCCT 9232
Oy 200 cagctgccccatgacgtgacgtgacattaaattcagaatactatcttcccaagg 259
Db 9231 GGGCTGGCGCGACCTCCACCTCGACACCCAGCCTGACAGCCACTGTCTGCCCAAGGT 9172
Oy 260 a 260
Db 9171 A 9171

RESULT 6
US-10-103-313-622
;; Sequence 622, Application US/10103313
;; GENERAL INFORMATION: a) al
;; APPLICANT: Rosen et al.

;; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;; FILE REFERENCE: P1207C1
;; CURRENT APPLICATION NUMBER: US/10/103,313
;; PRIOR APPLICATION REMOVED - See File Wrapper or Palm
;; NUMBER OF SEQ ID NOS: 653
;; SOFTWARE: PatentIn Ver. 2.0
;; SEQ ID NO 622
;; LENGTH: 1182
;; TYPE: DNA
;; ORGANISM: Homo sapiens
US-10-103-313-622

Query Match 11.5% Score 37; DB 6; Length 1182;
Best Local Similarity 50.3%; Pred. No. 0.03;
Matches 91; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

Oy 80 ggaagagattgaacgtgattgacgaacacgagccctgcatgcaagacagagagcca 139
Db 5 gcaagagctggagccctggtgtagggtagggagagccgcccgaagcctgagctatcgcggtg 64
Oy 140 catgacctacacagatgctggtgacgaggtccagagatgaccttgccttctcccccac 199
Db 65 cctgacctacgcaagctgctgctgagatccagcttccatcagctggtgctccct 124
Oy 200 cagctgccccatgacgtgacgtgacattaaattcagaatactatcttcccaagg 259
Db 125 ggggtgctgagcccttcccttccctgacacacccctgacacccatgctgcccagaaggt 184
Oy 260 a 260
Db 185 a 185

RESULT 7
US-10-105-299-7203/c
;; Sequence 7203, Application US/10105299
;; GENERAL INFORMATION: a) al
;; TITLE OF INVENTION: Human Secreted Proteins
;; FILE REFERENCE: PS950
;; CURRENT APPLICATION NUMBER: US/10/105,299
;; CURRENT FILING DATE: 2002-03-26
;; NUMBER OF SEQ ID NOS: 15197
;; PRIOR APPLICATION REMOVED - See File Wrapper or Palm
;; SOFTWARE: PatentIn Ver. 2.0
;; SEQ ID NO 7203
;; LENGTH: 1182
;; TYPE: DNA
;; ORGANISM: Homo sapiens
US-10-105-299-7203

Query Match 11.5% Score 37; DB 6; Length 1182;
Best Local Similarity 50.3%; Pred. No. 0.03; Indels 0; Gaps 0;
Matches 91; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

Oy 80 ggaagagattgaacgtgattgacgaacacgagccctgcatgcaagacagagagcca 139
Db 1178 GCAGGAGCTGGACCTCTGTGGTAGGGTGGAGGCGCCGCCCAAGCCTGGACTATCGCGTGTG 1119
Oy 140 catgacctacacagatgctggtgacgaggtccagagatgaccttgccttctcccccac 199
Db 1118 CTGCTCCCTAGCCGACGAGTGTGCTGGAGATCCAGTGTCTTCAAGCGTGGTGGCCCT 1059
Oy 200 cagctgccccatgacgtgacgtgacattaaattcagaatactatcttcccaagg 259
Db 1058 GGGGTGGCGCGACCTCCACCTCGACACCCAGCCTGACAGCCACTGTCTGCCCAAGGT 999
Oy 260 a 260
Db 998 A 998

RESULT 8
 US-10-106-698-892
 : Sequence 892, Application US/10106698
 : GENERAL INFORMATION:
 : APPLICANT: Ruben et al.
 : TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptides
 : CURRENT FILING DATE: 2002-03-27
 : CURRENT FILING DATE: 2002-03-27
 : PRIOR APPLICATION NUMBER: PCT/US00/26524
 : PRIOR FILING DATE: 2000-09-28
 : PRIOR APPLICATION NUMBER: US 60/157,137
 : PRIOR FILING DATE: 1999-09-29
 : PRIOR APPLICATION NUMBER: US 60/183,280
 : PRIOR FILING DATE: 1999-11-03
 : PRIOR FILING DATE: 1999-11-03
 : SOFTWARE: Patent In Ver. 3.0
 : SEQ ID NO 892
 : LENGTH: 988
 : TYPE: DNA
 : ORGANISM: Homo sapiens
 : FEATURE: misc_feature
 : NAME/KEY: equals a.t.g. or c
 : OTHER INFORMATION: equals a.t.g. or c
 : LOCATION: (977)..(977)
 : NAME/KEY: equals a.t.g. or c
 : OTHER INFORMATION: n equals a.t.g. or c
 : US-10-106-698-892
 Query Match 9.74; Score 31.4; DB 6; Length 988;
 Best Local Similarity 57.54; Pred. No. 1.4; 34; Indels 0;
 Matches 50; Conservative 3; Mismatches 0;
 QY 187 acctctctccacagctgccacagctgagctgacatgaattcagaactatc 246
 DB 752 atttctctccacagctgccacagctgagctgacatgaattcagaactatc 246
 QY 247 caattcccaagtgaaattgtttctcc 273
 DB 851 tyttgcggatgttaagtccttctcc 877
 RESULT 9
 US-09-975-254-18019
 : Sequence 18019, Application US/0975254
 : GENERAL INFORMATION:
 : APPLICANT: Beck, Gregory R.
 : APPLICANT: Byrum, Joseph R.
 : APPLICANT: Byrum, Gregory R.
 : APPLICANT: La Rosa, Thomas J.
 : TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
 : CURRENT FILING DATE: 2001-10-12
 : CURRENT FILING DATE: 2001-10-12
 : PRIOR APPLICATION NUMBER: US/09/263,191
 : PRIOR FILING DATE: 1999-03-05
 : NUMBER OF SEQ ID NOS: 31255
 : SEQ ID NO 18019
 : LENGTH: 254
 : TYPE: DNA
 : ORGANISM: Glycine max
 : FEATURE: misc_feature
 : NAME/KEY: equals a.t.g. or c
 : OTHER INFORMATION: Clone ID: 700955037H1
 : US-09-975-254-18019
 Query Match 9.54; Score 30.6; DB 5; Length 254;
 Best Local Similarity 53.84; Pred. No. 1.3; 54; Indels 0;
 Matches 63; Conservative 0; Mismatches 0;
 QY 187 acctctctccacagctgccacagctgagctgacatgaattcagaactatc 246
 DB 752 atttctctccacagctgccacagctgagctgacatgaattcagaactatc 246
 QY 247 caattcccaagtgaaattgtttctcc 273
 DB 851 tyttgcggatgttaagtccttctcc 877

Search completed: April 19, 2002, 11:07:06
Job time: 10746 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: April 19, 2002, 07:21:40 ; Search time 4168.4 Seconds
(without alignments)
6910.103 Million cell updates/sec

Title US-09-763-292-1
Seque.ce: 1 cttcaatgagctctttgtg.....acatattattaaataga 1746

Scoring table: IDENTITY_NOC
Gapop 10.0 , Gapext 1.0

Searched: 172140 seqs, 824859755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenBank:

1: gb.ba.*

2: gb.tcd.*

3: gb.tcd.*

4: gb.ov.*

5: gb.pat.*

6: gb.ph.*

7: gb.pl.*

8: gb.pr.*

9: gb.ro.*

10: gb.ro.*

11: gb.sts.*

12: gb.sts.*

13: gb.sts.*

14: gb.sts.*

15: gb.sts.*

16: gb.sts.*

17: gb.sts.*

18: gb.sts.*

19: gb.sts.*

20: gb.sts.*

21: gb.sts.*

22: gb.sts.*

23: gb.sts.*

24: gb.sts.*

25: gb.sts.*

26: gb.sts.*

27: gb.sts.*

28: gb.sts.*

29: gb.sts.*

30: gb.sts.*

31: gb.sts.*

32: gb.sts.*

33: gb.sts.*

34: gb.sts.*

35: gb.sts.*

36: gb.sts.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1739.5	99.5	1746	6	AR071575	AR071575 Sequence
2	1739.5	99.5	1746	6	HUMCYPC219	M61854 Human cytoc
3	1532.4	93.5	1669	6	E14931	E14931 Human mRNA
4	1534.5	97.9	1845	9	HUMCYPC2C9A	M61857 Human cytoc
5	1534.5	97.9	1845	6	AR071579	AR071579 Sequence
6	1529.8	87.5	1834	6	AR071576	AR071576 Sequence
7	1509.8	86.3	1843	6	E02279	E02279 Human cytoc
8	1509.8	86.3	1843	6	E02279	E02279 Human cytoc
9	1503.8	86.0	1826	9	HUMCYPH2	D00173 Homo sapien
10	1495.8	85.7	1814	9	S46963	S46963 putative CY
11	1458.5	81.5	1473	22	E10866	E10866 cDNA encodi
12	1456.4	83.4	2395	9	HUMSVF450A	L07093 Human cytoc
13	1436.4	82.3	1892	6	AR071581	AR071581 Sequence
14	1413.6	81.0	1444	6	E14930	E14930 Artificial
15	1367.6	78.3	1577	9	HUMCYPMPA	M21939 Human cytoc
16	1366.6	78.3	1576	9	HUMCYPMPC	M15331 Human liver
17	1343.4	76.9	1473	22	E10863	E10863 cDNA encodi
18	1343.4	76.9	1473	22	E10863	E10863 cDNA encodi
19	1264.5	72.4	2009	6	AR071577	M61856 Human cytoc
20	1264.5	72.4	2009	6	AR071577	AR071577 Sequence
21	1264.5	72.4	2258	6	AR071580	AR071580 Sequence
22	1264.5	72.4	2258	6	AR071580	M61853 Human cytoc
23	1238.8	71.0	1346	9	HSCF450	X65962 H.sapiens m
24	1183.2	57.8	1276	9	HUMCYPC2C17	M61858 Human cytoc
25	1175.6	57.3	1866	9	HSJIC2	Y00498 H. sapiens
26	1174.5	57.2	1901	9	SS3046	SS3046 cytochrome
27	1173.8	57.2	1473	22	E10866	E10866 cDNA encodi
28	1173.8	57.2	1473	22	E10866	E10866 cDNA encodi
29	1170.6	57.0	1829	9	HUMCYPAK	AR071578 Sequence
30	1142.2	65.4	1807	9	HUMCYPMPC	M21941 Human cytoc
31	1138.6	65.2	1441	9	HUMCYPMPC	M21940 Human cytoc
32	1121.6	64.2	1868	9	HUMCYPB	M17398 Human cytoc
33	1109.8	63.6	1473	22	E10639	E10639 Human cDNA
34	1109.8	63.6	1473	22	E10862	E10862 cDNA encodi
35	1106.6	63.4	1473	22	E10863	E10863 cDNA encodi
36	1105.5	63.3	1473	22	E10864	E10864 cDNA encodi
37	1097.8	62.9	1737	9	HUMCYPMPC	M21942 Human cytoc
38	1076.4	61.6	1737	9	PABCY205	M55664 Rabbit cyto
39	1076.4	61.6	1737	9	PABCY205	M55664 Rabbit cyto
40	1065.5	61.0	1662	4	PABCY45H23	D00190 Oryctolagus
41	1051.2	60.8	1827	4	PABCY4503	J02716 Rabbit cyto
42	1055.2	60.4	1871	4	PABCY4501	M11299 Rabbit live
43	1050.6	60.1	1679	4	AF016248	AF016248 Canis fam
44	1038.6	59.5	1758	4	SSC15D09	Z93100 S.scrofa CY
45	1038.2	59.5	1808	10	HANP4502	D11435 Mesocricetu

ALIGNMENTS

RESULT 1					
AR071575	AP071575	1746 bp	DNA	PAT	18-FEB-2000
LOCUS	Sequence 2	from patent US 5912120.			
DEFINITION	AR071575				
ACCESSION	AR071575				
VERSION	AP071575.1	GI:7222463			
KEYWORDS	Unknown.				
SOURCE	Unknown.				
ORGANISM	Unknown.				
REFERENCE	Unpublished				
Unpublished	(accession to 1746)				
AUTHORS	Goldstein, J.A. and De Morais, S.M.F.				
TITLE	Cloning, expression and diagnosis of human cytochrome P450 2C19:				
JOURNAL	the principal determinant of s-mephenytoin metabolism				
FEATURES	Patent: US 5912120-A 2 15-JUN-1999:				
source	Location/Qualifiers				
	1..1746				
BASE COUNT	477 a 418 c 365 g 486 t				
ORIGIN	/organism="unknown"				

Query Match	99.58%	Score 1739.5	DB 6	Length 1746	
Best Local Similarity	99.84%	Score 1746.0	DB 6	Length 1746	
Matches 1742	Conservative	Mismatches	Indels	Gaps	0
py	1	cttcaatggatccttttggctgtgctgtgctgtctctctcattgtctgtctctcttca	60		
pb	1	CTTCAMATGATCTTTTGGCTGTGTCTGTCTCTCTCTCTCTCTCTCTCTCTCA	60		
py	61	tctggagcagagctctggaggggaaaaactcctcttggccctcaactcctctccacga	120		
pb	61	TCTGGAGCAGAGCTCTGGAGGGGAAAACTCCTCTTGGCCCTCAACTCCTCCAG	120		
py	121	tctggaaatcctcagatagatcaagaatcgacgaactcttaacaaactctcaca	180		
pb	121	TTGGAAATCTCCTACGATAGATATTAAGGATGTGACGAACAATCTTAACCAATCTCTCAA	180		
py	181	aaatcctagggcctgtgttcactctgtatttggcctggacgaatggtgctgcgatg	240		
pb	181	AAATCTTAGGCCCTGTGTTCACCTGTATTATTGGCCTGGACGAATGGTGTGCTGCAATG	240		
py	241	gatatgaagctgctgagagaaacccctgactgctctggagagactctctggagagcc	300		
pb	241	GATATCAAGTGGTGAAGGAAGCCCTGATGATCTTCGAGAGAGTGTTCGGAAGAGGCC	300		
py	301	atttccctcctggctgaaagcgaacagagatttggaaactgttttcagcaatgagaaga	360		
pb	301	ATTTTCCCACTGGCTGAAGAGCTACAGAGATTTGGAAATCGTTTACGCAATGGAAGA	360		
py	361	gatggaagagatcgagcttttctcctcctcagcgtcggaaatttggatggagga	420		
pb	361	GATGGAAGAGATCGAGCTTTTCTCCTGATGAGCTCGGAAATTTGGGAAGGAGGA	420		
py	421	ggagcatggagacccgtgttcaagagagagccgcctgactctggagagcttgagaaa	480		
pb	421	GGAGCAATGAGAGACCGTGTTCAGAGAGAGCCCGCTGCTGTGGAGAGTGTGAGAAAA	480		
py	481	caaagctcaactcgttgatcccaatttcttggtgctgctcctgcgatcgatct	540		
pb	481	CAAGAGCTTCACTGATTCGCAATTTCTTGGGCTGTGTCTCTCAATGATCT	540		
py	541	gctcctatttccagaacgttctgatataaagatcncagactcttaacttgatgg	600		
pb	541	GCTCCATATTTCAGAAAGCTTTCGATTATAAGATCAGCAATTCCTTAAGTGTGG	600		
py	601	aaaaatgtaatgaacaatcaggattgtaagcaccctggatcagatatagcaatt	660		
pb	601	AAAAATGTAATGAACAATCAGGATTGTAAGCACCCTGGATCAGATATAGCAATT	660		
py	661	tcccaactcattgatatttccggagagaccatcaaaactctgaactaaactcgtct	720		
pb	661	TCCCAACTCATTGATTTCCGGAGAGACCATCAAACTCTGAACCTAAACTCAGT	720		
py	721	ttaaggaaatgataatttggagaagaaagaaacacaaagatcgatgacatcaaca	780		
pb	721	TTAAGGAATGATAATTTGGAGAAGAAAGAAACACCAAGACATCGATGACATCAACA	780		
py	781	acctcgggaacttatgatgtctcctgatacaaatggagaaagaaagcaaaccaac	840		
pb	781	ACCTCGGAACCTATTGATATTCCTCTGATCAAAATGGAGGAAGAAAGCAACCAAC	840		
py	841	agctcgaatctactatgaaacttgataactcagcagctgacttaacttgagatgaga	900		
pb	841	AGCTCGAATCTACTATGAAACTTGGTAATCATCTCAGCTGACTTACTTGGAGTGGGA	900		
py	901	cagagcaacaagcacaacctgagatagctctctctctctctgtaagcaccacgaag	960		
pb	901	CAGAGCAACAAGCACAACCTGAGATAGCTCTCTCTCTCTCTGTAAGCAACCAAGAG	960		
py	961	tcacagctaaactccagagaaatgaacgctgtctgctgacgacagacagccctgca	1020		
pb	961	TCAAGCTAAAGTCCAGAGAGATTGACGTGTCTATTCGCAACAACCGAGAGCCCTGCA	1020		

[illegible]


```

Db 1321 GCTGCGCCGATCGAGCTGTTTATTCCTGACCTTCATTTACAGCACTTTAACTCGA 1380
Oy 1381 aactctgagtgacccaagagaccttcaacaactcctggttcaatgagattgctctg 1440
Db 1381 AATCTCTGATTCGACCAAGAGACCTTGACCAACTCTCTGTGTCAATGAAITGCTTGT 1440
Oy 1441 tccgcgcctctatcagctggtcttctcattcctctctgaaagacacagatgctgctg 1500
Db 1441 TCCGCCCTTCTACAGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1500
Oy 1501 ctctgagctgctccctgagctctctctctctctctctctctctctctctctctct 1560
Db 1501 CTCTGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1560
Oy 1561 tctctgacgcctcatctcatctctctctctctctctctctctctctctctctctct 1620
Db 1561 TCTCTGACCGCTCATCTCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1620
Oy 1621 catcaaaagtttctctgagctgagctgagctgagctgagctgagctgagctgagct 1680
Db 1621 CATTAAAAAGTTCTCTGTCGAATATCTCTGTCATCTCTCTCTCTCTCTCTCTCTCTCT 1680
Oy 1681 atlgagtgccacaatactgctgactgctgactgctgactgctgactgctgactgact 1740
Db 1681 ATTGAGTGCACAAATGCTGACTGCTGACTGCTGACTGCTGACTGCTGACTGCTGACT 1740
Oy 1741 aataga 1746
Db 1741 AATAGA 1746

RESULT 3
E14931
LOCUS Human mRNA for hepatic cytochrome P450 2C19. 28-JUL-1999
DEFINITION
ACCESSION E14931
VERSION 1 GI:5709614
AUTHORS J. A. Schuster, J. A. Schuster, J. A. Schuster
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE
AUTHORS Baba, T., Kirita, S. and Aoyama, J.
TITLE MASS EXPRESSION SYSTEM OF MODIFIED SUBSTANCE OF CYTOCHROME P450 2C19
JOURNAL IN ESCHERICHIA COLI
COMMENT Parent: JP 1998033166-A 2 10-FEB-1998
SR10001 & CO LTD
JP 1998033166-A/2
PD 10-FEB-1998
PF 23-JUL-1996 JP 1996193015
PI BABA TAKAHITO, KIRITA SHIRO, AOYAMA JUNKO
PC C12N9/02, C07H1/04, C07K1/16, C12N1/21, C12N15/09, C12N5/02, PC
C12H1/19,
CC (C12N1/21, C12H1/19):
CC strandedness: Double;
CC topology: Linear;
FH Key Location/Qualifiers
FT source 1..1669
FT /organism="Homo sapiens"
FT /tissue_type="liver"
FT CDS
FT /product="hepatic cytochrome P450 2C19"
FT /locus_tag="P450_2C19"
FT /db_xref="taxon:9606"
FEATURES
source
BASE COUNT
ORIGIN
449 a 405 c 363 g 452 t

```

```

Query Match 93.5% Score 1632.4 DB 6 Length 1669
Best Local Similarity 99.5% Pred. No. 0
Matches 1636 Conservative 0 Mismatches 6 Indels 0 Gaps 0
Oy 1 ctccaagagatccttttgagctcttgctgctctctctctctctctctctctctctctca 60
Db 20 CTTCATGATGCTTTTGAGCTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCA 87
Oy 61 tctgagagacagctctgagagagagagagagagagagagagagagagagagagagag 120
Db 61 TCTGAGACAGAGCTCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
Oy 88 TCTGGAGACAGAGCTCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 147
Db 88 TCTGGAGACAGAGCTCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 147
Oy 121 ttgaaatattctacagatagatatlaagagatgcagcaaatcttaaccaatctctca 180
Db 121 TTGAAATATCTACAGATAGATATAGAGATGTCAGCAAACTCTTACCACTCTCTCA 207
Oy 181 aaactatagctgctgctcaactctctctctctctctctctctctctctctctctct 240
Db 181 AATCTATGCTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 240
Oy 241 gatatgaagtgagagagagagagagagagagagagagagagagagagagagagag 300
Db 241 GATATGAAGTGGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300
Oy 268 GATATGAAGTGGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 327
Db 268 GATATGAAGTGGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 327
Oy 301 atttccactggttgaagagagctcaacagagattggagctctttccagcaatgaga 360
Db 301 ATTTCCTGCTGCTGAGAGAGCTTACAGAGAGATTTGGAACTGCTTTCACCATGCA 387
Oy 361 gatgagagagattcgcccttctctctctctctctctctctctctctctctctct 420
Db 361 GATGAGAGAGATTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420
Oy 388 GATGCAAGGAGATCCGGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 447
Db 388 GATGCAAGGAGATCCGGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 447
Oy 421 gagcattgaagagacgtcttcaagagagagagagagagagagagagagagagagag 480
Db 421 GAGCATGAGAGAGCTCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 507
Oy 481 ccaagctctacccctgagctctctctctctctctctctctctctctctctctctct 540
Db 481 CCAAGCTCTACCCCTGAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 540
Oy 541 gctccattatttccagagagcttctctctctctctctctctctctctctctctctct 600
Db 541 GCTCCATTATTTCCAGAGAGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 627
Oy 568 GCTCCATTATTTCCAGAGAGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 627
Db 568 GCTCCATTATTTCCAGAGAGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 627
Oy 601 aaaaattgaatgaagacacagagattgaagacacacacacacacacacacacacac 660
Db 601 AAAAAATGAATGAAGACATCAGATTTGTAAGACCCCTTGATCCAGATATGCAATATT 687
Oy 661 ttcacctatcattgatttctctctctctctctctctctctctctctctctctctct 720
Db 661 TTCCACTATCATTTGATTATTTCCCGGAGCCCATACAAATATCTTAAACACCTTGCT 747
Oy 721 ttatggaagtgatatttggagagagagagagagagagagagagagagagagagagag 780
Db 721 TTATGGAAGTGATATTTGGAGAGAGTAAAGACACCAAGATCTGATGATGATGATG 807
Oy 781 acctctgagagcttctctctctctctctctctctctctctctctctctctctct 840
Db 781 ACCCTCTGGAGCTTTATGATTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 867
Oy 841 agtctgaattcactattgaaacacacacacacacacacacacacacacacacacac 900
Db 841 AGTCTGAATTCATTTGAAACCTTGGTAATCTCTGAGCTGACTTACTTGGAGCTGG 927
Oy 901 cagagacacagacacacacacacacacacacacacacacacacacacacacacacag 960
Db 901 CAGACACACAGACACACCTTGAGATAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 987
Oy 961 tcacagctaaagctcagagagagagatgaacagagagagagagagagagagagagag 1020
Db 961 TCACAGCTAAAGCTCAGAGAGAGATGACGTTGCTGTTGGAGAGAGAGAGAGAGAG 1047

```


Db 728 TATGAAAGTTATATTTGGAAAGTAAAGAACACACGANGAATCAATGACATGAACA 787
Oy 781 accctcggagcttattgattgcttctctgatacaaatggagaagaaacaaacacac 840
Db 788 ACCCTGAGAGCTTATGATGCTTCTGATGAATGGAAGAAAGAACACACACAC 847
Oy 841 agctcgaattcattgataaacttgatcaactcagctgagcttacttggagctgga 900
Db 848 CATCTGAAATTTACTATTGAAAGTTGGAAAGACTGCAATGACTTGTGGAGCTGGCA 907
Oy 901 cagagcaacaaagacacccctgagatgctctctctctctgctgaagcaaccagagg 960
Db 908 CAGAGAGCAAGCAACACCTGAGATATGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 967
Oy 961 tcaagctcgaagctcaggaagagattgacgctgctatggcagaaacggagccctgca 1020
Db 968 TCAAGCTTAAGGTCAGAGAGATGAGCTGATGAGCAAGAACCCGAGCCCTCTGCA 1027
Oy 1021 tgcagagagagggccacacacccacagatgctggtgacagagtcagagataca 1080
Db 1028 TGCAG 1087
Oy 1081 tgaagaaagtaacttctgctctctctctctctctctctctctctctctctctct 1140
Db 1088 TGAAGATTTCCCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1147
Oy 1141 acctcattcccaagggagcaacatataacttccctcactctctgctacagagaa 1200
Db 1148 ATCTCATTTCCCAAGGGCAGACACCATATTAATTCCTGACTCTCTGCTACATCA 1207
Oy 1201 aagaattttcccaacccagagagattgagacccctgacactttctgagagtggaatt 1260
Db 1208 AAGATTTTCCCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1267
Oy 1261 taaagaaagtaacttctgctctctctctctctctctctctctctctctctctct 1320
Db 1268 TTAAGAAAGTAAATCTCAATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1327
Oy 1321 gcttgagcccgatgagctgttttattctcagacttcttctacagactttaactga 1380
Db 1328 CCTGCGCGGAGTGGAGCTGTTTATTTCTGACTTCCATTTTACAGACATTTAAGCTGA 1387
Oy 1381 aatctctgattgacccaagagacttggacacactctctgttgcattgatttctctg 1440
Db 1388 ATCTCTCTGCTGAGCCCAAGAGAGCTTGAACACCTGACCTGCTGCTGCTGCTGCT 1447
Oy 1441 cccagcccttctatcagctgctctctctctctctctctctctctctctctctct 1500
Db 1448 TGGCGGCTCTCTACGAGCTGTGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1507
Oy 1501 ctctgtgctgctcctgagctctctctctctctctctctctctctctctctct 1553
Db 1508 CTGCTGTGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1567
Oy 1554 gtagcttcttgcacccgactcactcacttctctctctctctctctctctctctct 1613
Db 1568 TAATGCTCTTCTCACTGCTCACTGCTCACTGCTCACTGCTCACTGCTCACTGCT 1627
Oy 1614 cagctccactta-----aaaaagtttccagtggaatatattctgctattccca 1653
Db 1528 CGAGCTTCTATACGAGAGTTTCTCTATGTTTCTCTGTTGCAATATATCTGCTATCTCA 1697
Oy 1664 tacttataatggttactatgagtcacatgctgactgactgactgactgactgactgact 1723
Db 1688 TACTCTGTACAGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1747
Oy 1724 taaactattattataataga 1746
Db 1748 TTAATATGTTATTATAATAGA 1770

RESULT 8

E02279
LOCUS Human liver cytochrome P450MP gene. PAT 29-SEP-1997
DEFINITION Human liver cytochrome P450MP gene.
ACCESSION E02279
VERSION E02279.1 G1:2170516
KEYWORDS JP 1990072879-A/1
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
REFERENCE 1 (bases 1 to 1843)
AUTHORS Ogita.S., Jinriki.N., Fujitani.T., Kamakaki.T., Komori.M. and Miura.T.
TITLE RECOMBINANT PLASMID INTEGRATED WITH GENE OF HUMAN LIVER CYTOCHROME P- 450MP, YEAST PRODUCING P-450MP AND PRODUCTION THEREOF
JOURNAL Patent: JP 1990072879-A 1 13-MAR-1990;
COMMENT AGENCY OF IND SCIENCE & TECHNOL
PN JP 1990072879-A/1
PF 09-SEP-1988 JP 1988225955
PI OGITA SATORU, JINRIKI NARIKO, FUJITANI TOMOMICHI, P1 KAMAKAKI TETSUYA
PI KOMORI MASAYUKI, MIURA TOSHIAKI
PC C12N15/53.C12N1/16.C12N15/81.(C12N1/16.C12R1:865); CC
STRANDEDNESS: Double;
CC topology: linear;
CC topological no.;
CC anti-sense: No;
CC *source: tissue_type=liverr;
CC *source: clone-phiA6;
FH Key Location/Qualifiers
FH 5'UTR 1..4
FT CDS 5..1471
FT FT_peptide 5..1471/product='human liver cytochrome P450MP' FT
FT 3'UTR 1472..1843.
FEATURES
source
1..1843
/organism='Homo sapiens'
/db_xref='taxon:9606'
BASE COUNT 512 a 424 c 377 g 530 t
Query Match 85.1%; Score 1506.4; DB 6; Length 1843;
Best Local Similarity 92.1%; Pred. No. 0;
Matches 1626; Conservative 0; Mismatches 111; Indels 23; Gaps 3;
Oy 4 caatggtccttttgatccttgctgctgctgctgctgctgctgctgctgctgctgct 63
Db 3 CTATGATGCTACTTGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 62
Oy 64 ggaagagagctctggaag 123
Db 63 GGAGACAGAGCTCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 122
Oy 124 gaatatctcagatagattatgaagagagagagagagagagagagagagagagagag 183
Db 123 GAATATCTCTAGATAGATATTAAGACATACAGAAATCTTTAACCATCTCTCAAGG 182
Oy 184 tctatgctcctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 243
Db 183 TCTATGCGCGGTCTCTACCTCTGTTATTTGGCTGAAACCATAGTGTGCTGCTGATG 242
Oy 244 atgaagtgag 303
Db 243 ATGAAGCAGTGAAGAT 302
Oy 304 tccactgctgag 363

```
Db 303 TCCACTGCGTAAGAGCTAACAGAGATTGGAAATGTGTTTCAGCAATGGAAAGAAAT 362
Oy 364 ggaagagatggcgcttccctccctcagcgtcggaattttggaatgggagagga 423
Db 365 GAAAGAGATCCGGGCTTCTCCCTCATGAGCTCGCGAATTTGGGATGGGAGAGGA 422
Oy 424 gcatggagacgcgtgtcgaagagaaccccgcttgcgtggaagattgagaagaaaca 483
Db 425 gcatggagacgcgtgtcgaagagaaccccgcttgcgtggaagattgagaagaaaca 482
Oy 484 aggcctcacccttgatcccaacttccatctgggctgctccctgcaatgatctgct 543
Db 485 aggcctcacccttgatcccaacttccatctgggctgctccctgcaatgatctgct 542
Oy 544 ccatatttccgaagacgtttcgatataagaatcacgaatttcttaacttgagaaa 603
Db 545 ccatatttccgaagacgtttcgatataagaatcacgaatttcttaacttgagaaa 602
Oy 604 aattgataaacaacatcaggatgtgaagcaccctcgatccagatcagatataatttt 663
Db 605 aattgataaacaacatcaggatgtgaagcaccctcgatccagatcagatataatttt 662
Oy 664 ccaactcatgattattcccggaagaccatacaaaatcttaaaacacttgcttta 723
Db 665 ccaactcatgattattcccggaagaccatacaaaatcttaaaacacttgcttta 722
Oy 724 ggaagatgattttgagaagataaagaacaaacgaagaatgatggatcaacaacc 783
Db 725 ggaagatgattttgagaagataaagaacaaacgaagaatgatggatcaacaacc 782
Oy 784 ctgggacttattgatctctctgatacaaaatggaagagaagcaaaccaacagt 843
Db 785 ctgggacttattgatctctctgatacaaaatggaagagaagcaaaccaacagt 842
Oy 844 ctgaattcactattgaaacttgtaactcactgcagctgacttacttgagctggagag 903
Db 845 ctgaattcactattgaaacttgtaactcactgcagctgacttacttgagctggagag 902
Oy 904 ctgatttacttgaagacttggaacactgcagctgacttacttgagctggagag 963
Db 905 ctgatttacttgaagacttggaacactgcagctgacttacttgagctggagag 962
Oy 964 cagctcgaagtcgaagagatgagcgtgtcattggcagaacacgcgacccctgcacg 1023
Db 965 cagctcgaagtcgaagagatgagcgtgtcattggcagaacacgcgacccctgcacg 1022
Oy 1024 acgacaggggcccacatgccctacacagatgctggtgacagaggtccagagatacatc 1083
Db 1025 acgacaggggcccacatgccctacacagatgctggtgacagaggtccagagatacatc 1082
Oy 1084 acctctcccaacacccctgccatgcagtgacctgacgttaaatctcagaacatcc 1143
Db 1085 acctctcccaacacccctgccatgcagtgacctgacgttaaatctcagaacatcc 1142
Oy 1144 tcattcccaaggggcaaccattatcaactccctcactctctgctacatcaacaaga 1203
Db 1145 tcattcccaaggggcaaccattatcaactccctcactctctgctacatcaacaaga 1202
Oy 1197 tcaattcccaaggggcaaccattatcaactccctcactctctgctacatcaacaaga 1253
Db 1198 tcaattcccaaggggcaaccattatcaactccctcactctctgctacatcaacaaga 1252
Oy 1254 agaaagttaactcactcctctccagcagaacacgalttggtggagagggcc 1323
Db 1255 agaaagttaactcactcctctccagcagaacacgalttggtggagagggcc 1322
Oy 1256 agaaagttaactcactcctctccagcagaacacgalttggtggagagggcc 1315
Db 1257 agaaagttaactcactcctctccagcagaacacgalttggtggagagggcc 1314
Oy 1324 tggcccgatggagctgtttttattcctgaccttccattttacagacatttaactgaaat 1383
Db 1325 tggcccgatggagctgtttttattcctgaccttccattttacagacatttaactgaaat 1382
Oy 1384 ctctgattgaaccaaagacgttgacacaaactgtgttgatgaatgatttctgctacc 1443
Db 1385 ctctgattgaaccaaagacgttgacacaaactgtgttgatgaatgatttctgctacc 1442
Oy 1377 ctctggttgaccacaagaccttgacacacacttcactggttcattgatttgcctctgtgc 1435
Db 1378 ctctggttgaccacaagaccttgacacacacttcactggttcattgatttgcctctgtgc 1434
```


hydroxylase activity
 Biochem. Int. 27 (5), 1073-1081 (1992)
 93075249
 Genbank staff at the National Library of Medicine created this
 entry [NCBI gibbs 11/17/92] from the original journal article.
 REMARK
 Author also gives sequence for another cytochrome P-450 that
 contains a 6-base deletion of nucleotides 783-788.
 COMMENT
 Location/Qualifiers
 1. .1814
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 1. .1814
 /gene="putative CYP2C9 gene"
 3. .1436
 /gene="putative CYP2C9 gene"
 /codon_start=-1
 /product="cytochrome P-450"
 /protein_id="BAB23864.2"
 /db_xref="GI:5705937"
 /translation="MILLSLRSSGRCKLPKPTPLPVGNILOIGIKDISKLNL
 SKYGVETLYGLKPIVYLVHGVKALVLEADLEEFSGRIPLAEPNRFGLVES
 NGKMKREIRFLSLMTNFGKRSIEDRYOEARKLVEELPKASPDITFLUGCA
 PLYVCSIFERRDYKQDFLNLEKLNKILSLPMIQCINNFSPIDYFFGTIN
 KLLANVMSKIDERYKROESDMNNPDYDFCLPMKREKRNQSPSEFTESLN
 TDAVYRVQDTSTLRTALLLKHPEYATVQETLIERVIGRNPSCHDRSIMPY
 TDAVYRVQDTSTLRTALLLKHPEYATVQETLIERVIGRNPSCHDRSIMPY
 PENLDTTPVYNGFASVPPTDGLFIPV"
 BASE COUNT 517 a 413 c 369 g 515 t
 ORIGIN

Query Match 85.7% Score 1495.8; DB 9; Length 1814;
 Best Local Similarity 52.8%; Pred. No. 0;
 Matches 1597; Conservative 0; Mismatches 107; Indels 17; Gaps 2;

Oy 43 gttctctctcttcaatcgtgacacagctctgagagagaaactccctcagcc 102
 Db 1 GTTTGCTCTCTCTTCACTCTGGACACAGAGCTCTGGAGAGGAAACTCCTCTCGCC 60

Oy 103 ccaactctctccagtgatggaatactctacagatagattaaaggaagtcaagaat 162
 Db 61 CCACCTCTCTCCAGGATGGAAATATCTACAGATAGGTATTAGGACATCAGCAAT 120

Oy 163 ccttaacaaatctcaaaaactatgtgacctgttcaactctgatttggcctgaac 222
 Db 121 CTTTAAACAATCTCTCAAGGCTCTATGGCCCTGTCTCACTCTGTCTTTGGCCTGAAC 180

Oy 223 gcatgggtgctgctgatgataagagtgtaggaagacccctgatctctggagag 282
 Db 181 CCATAGTGTGCTGCATGGATATGAAGCAGTGAAGGAGGCCCTGATTCTTGGAGAG 240

Oy 283 agtttctggagagccatttcccactggctgaagagcgtacacagagatttggaaac 342
 Db 241 AGTTTCTGGAGAGGACATTTCCCACTGGCTGAAGAGCTTACAGAGATTTGGAAATG 300

Oy 343 ttttcgaatggaagagatgaagagattcgcgtttctccctcagcagtcagaga 402
 Db 301 TTTTCAGCAATGGAAGAAATGGAAGGAGATCCGGCTTTCTCCCTCATGAGCTCGCGA 360

Oy 403 atttggagtggaagagagcattgagagcgtgttcaagagagagccctgccttg 462
 Db 361 ATTTTGGGATGGGAGAGGAGGATGAGAGCCGTGTTCAGAGAGAGCCCGCTGCTTG 420

Oy 463 tgaagagattggaagaaacaaagcttcaacctgtatccactttcattcctggctg 522
 Db 421 TGGAGAGTTTGAAGAAACCAAGGCTTCACCTGTATCCCACTTTCATCTCTGGGCTGTG 480

Oy 523 ctccctcgaatgctgctcctcattttccgaagacgtttcgattataataacacg 582
 Db 481 CTCCCTGCAATGATCTGCTCCATTATTTCCATAACGTTTTCATTATTAAGTACAG 540

Oy 583 aattcttaactgaggaataatgaatgaataacatcaggatgtgaagcaccctcaga 642
 Db 541 AATTTCTTAACCTAATGAAGAGTTGAATGAAGAACATCAAGATTTTGAGCAGCCCTCGA 600

Oy 643 tccagataatgaataatcttccatctcattgatttatttccctggagaccataaacaat 702
 Db 601 TCCAGATCTGCAATATTTTCTCATCATGATTACTTCCGGAGATCTACACCAAT 660

Oy 703 tacttaaaacctctctttatggaagatgattttgggaagagtaaaagacacccaag 762
 Db 661 TACTTAAAACGTTTCTTTATGAAAAGTTATTTTGGAAAAGTAAAGAACACCAAG 720

Oy 763 aatcagtgacatcaacaacccctcggaactttatgattgcttctctcgtacaaatggaga 822
 Db 721 ATCAATGGACATGAACACCCCTCAGAGCTTTATTCATGCTTCTCTGATGAATGGAGA 780

Oy 823 aggaagacaaacaaacagctgattcattatgaaaacttggtaactcactgcagctg 882
 Db 781 AGGAAGAACACAAACCACTGAAATTTCTATTGAAAGCTTGGAAACATCGCAAGTTG 840

Oy 883 actactgtgagctggagacacacacacacacacacacacacacacacacacacac 942
 Db 841 ACTTGTGTGAGCTGGGACAGACAGACAGACACACACACACCTTGAGATGCTCTCTCTCC 900

Oy 943 tctgaagac 1002
 Db 901 TCTCAGACACCCAGAGGTCAACAGCTTAAGTCCAGGAGAGATTGAACGTGTGATTGGCA 960

Oy 1003 gaacacagacccctctgctacacagacagagagacacacacacacacacacacacac 1062
 Db 961 GAAACCGAGCCCTCGCATGACAGACAGACGACACACATGCTCTACACAGATGCTGTGTTGC 1020

Oy 1063 acgaggtccagagatacactcgaactcattccacacacacacacacacacacacacacac 1122
 Db 1021 AGAGGTCCAGAGATACATGACCTTCTCCACACAGCTGCCCCCTGAGTGTG 1080

Oy 1123 acgttaactcagaactcactcattcccaaggacacacacacacacacacacacacacac 1182
 Db 1081 ACATTAATTTGAAACATCTCATCTCCAGGACACACACATTAATTTCCCTGACCTT 1140

Oy 1183 clgtgctacatgac 1242
 Db 1141 CTGTGCTACATGACACAAAGAAATTTCCCAACCCAGAGATGTTTGACCTCATCATCTTC 1200

Oy 1243 tggatgaagtggaatcttctccacacacacacacacacacacacacacacacacacacac 1302
 Db 1201 TGAATGAGGTGGCAATTTTAAAGAAAGTAATACTTCTGCTTCTCAGAGGAAAC 1260

Oy 1303 gaatttctggaagagac 1362
 Db 1261 GGAATTTGTGTGGAGAGAGCCCTGGCGGCAATGGAGCTGTTTTTATTCCTGACCTCATTT 1320

Oy 1363 tacaaaacttaacctgaaactctctgattgacccaaagacacacacacacacacacacacac 1422
 Db 1321 TACAAACTTTAACTGAAATCTCTGTTGACCCAAAGAACCTTGACACACCTCCAGTTG 1380

Oy 1423 tcaatgaattgcttctgctccac 1482
 Db 1381 TCATGGATTTGGCTCTGTGCGCCCTCTACAGCTGCTCTCATTTCTCTGTGAGAA 1440

Oy 1483 gcacaaatgagctggctcctcctgctgctcctcctcctcctcctcctcctcctcctcctc 1535
 Db 1441 GACAGATGGCTGGCTGCTGTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1500

Oy 1536 tccaaatctcaactatctgatacttctctgacccctcactcacaacttttcccttccccc 1595
 Db 1501 TCAATCTTCACTATCTGATGCTTTTTCACCTGTGCTCATCTCAGATTTTCCCTTCCTT 1560

Oy 1596 caagctcagagaaactcagctcactta-----aaaaagttcacttgcaaa 1645
 Db 1561 GAAGATCTAGTGAACATTCGACCTCCATAGGAGAGGTTTCTCTATGTTTCACTGTGCAAA 1620

Db	1141	ATTTCCAAAGGGCACACCATATTAACCTCCCTCACTCTGCTACATGACACAGAA	1200	PFSLRLKDDXFLNKLNLNLRIVSTPWIOJNNPPTIIDYFPTGTHNKLKLIAM ESDLEKVEKESDNNPDRDFIDCLFKMEKONQOSEFTIENLVITAAUJLGA TETISITLYALLLLKHPVTAKEVEIERVWGRNRSPCMQORHPMYDAVHEVO RYIDILPTSLPHATVCDVAFRNLIPKGTITLTSVLHDKNKEFPNPFDRHELD EGGNFKSNFPEFSAKRCVCEGLARWELFLFIFLQNFNLKSLIDPKDLOITPV VNGFASVPPFIOLUFIPIV
Qy	1206	tttcccaaccagagatgttgacccctgctacattctctgagtagaggtgaaattttaaag	1265	
Db	1201	TTCCCCACCCAGAGATGTTGACCCCTGCTCACTTCTCGATGAAGGTGAATTTAAG	1260	
Qy	1266	aaaagtaactactctgcttcttcagcaggaaacagagattgtgtggagagggctg	1325	
Db	1261	AAAGTACTACTTCTGCTTCTCCAGAGAACGAGATTTGTGGAGAGGGCTG	1320	
Qy	1326	gccgcagtagagctgttttcttcagcactctattctacagactttaacctgaaact	1385	
Db	1321	GCCCGCATGGAGCTGTTTTTCTCTGAGCTTCATTTTACAGAACTTTAACCTGAATCT	1380	
Qy	1386	ctgatgaccacaagacttgacacactctgtgtgcaatggattgtgtctgccg	1445	
Db	1381	CTGATGTACCCAAAGGACCTTGACACACTCTGTGTGTCATGATTTGCTCTGTCGG	1440	
Qy	1446	ccctctctacactctgct	1478	
Db	1441	CCCTCTCACTGCTGCT	1473	
RESULT	12			
LOCUS	HUMSVP450A	2395 bp	mRNA	25-JUL-1993
DEFINITION	human cytochrome P450C18 and P450C19 splice variant or composite clone mRNA.			
ACCESSION	L07093			
VERSION	L07093.1	GI:307443		
KEYWORDS	Homo sapiens cDNA to mRNA.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
AUTHORS	I. (bases 1 to 2395) Gollard, W., Faletto, M.B., Blaisdell, J.A., Raucy, J.L. and Gollard, W.			
TITLE	Correction: Cloning and expression of complementary cDNAs for multiple members of the human cytochrome P45011C subfamily			
JOURNAL	Biochemistry 32, 1390-1390 (1993)			
MEDLINE	93192243			
COMMENT	Clone 254c was originally listed as a separate gene under Accession Number M61858. Resequencing demonstrated that it is not a separate gene, but is a splice variant or a composite cDNA. The 5' portion corresponds to a partial 2C18 clone, and the 3' portion corresponds to a partial 2C19 clone. The 5' end from 1 to 931 corresponds to a partial allelic variant of 2C18 beginning at base 361 of 2C18 continuing through base pair 1291 marking the end of exon 8. There are two allelic differences from our previously reported 2C18 clone 29C(M61855), a G to A change at 387 coding for Ile 129 instead of Met 129 and a G to A change at bp 1004 coding for Gln335 instead of Arg. The 3' end of the clone, from bases 932-2395 corresponds to a partial 2C19 clone (M61857) beginning at position 189 of 2C19 (beginning of exon 2) through to position 1562. There is an allelic variation from the previously reported 2C19 sequence, an A to G substitution at 991 of 2C19 coding for Gln335 instead of Arg.			
FEATURES	Location/Qualifiers			
SOURCE	1. 2395 /organism="Homo sapiens" /db_xref="taxon:9606"			
CDS	80..2236 /note="splice variant" /codon_start=1 /product="cytochrome P450" /protein_id="AAA36660.1" /db_xref="GI:307444"			
	LYFLPMVVLHGVTEVKALDLGEEFSGRGHFPLAEAHGFGIVFSNKKRKEIPVFTI FORFDYKDXFLNKLNLNLRIVSTPWIOJNNPPTIIDYFPTGTHNKLKLIAM ESDLEKVEKESDNNPDRDFIDCLFKMEKONQOSEFTIENLVITAAUJLGA TETISITLYALLLLKHPVTAKEVEIERVWGRNRSPCMQORHPMYDAVHEVO RYIDILPTSLPHATVCDVAFRNLIPKGTITLTSVLHDKNKEFPNPFDRHELD EGGNFKSNFPEFSAKRCVCEGLARWELFLFIFLQNFNLKSLIDPKDLOITPV VNGFASVPPFIOLUFIPIV			

[illegible]

RESULT	2:1
AR071581	
LOCUS	1892 bp DNA
DEFINITION	Sequence 14 from patent US 5912120.
AR071581	
ACCESSION	AR071581.1 GI:7222469
VERSION	
KEYWORDS	
SOURCE	Unknown.
ORGANISM	Unclassified,
REFERENCE	1 (bases 1 to 1892)
AUTHORS	Goldstein,J.A. and De Morales,S.M.F.
TITLE	Cloning, expression and diagnosis of human cytochrome P450 2D19:
JOURNAL	the principal determinant of S-mephenytoin metabolism
FEATURES	Patent: US 5912120-A 14 15-JUN-1999;
source	Location/Qualifiers /organism="unknown"
BASE COUNT	481 a 396 c 358 g 500 t 157 others
ORIGIN	

```

Query Match      82.3% Score 1436.4 DB 6: Length 1892:
Best Local Similarity 86.5% Ref. No. 0:
Matches 1525; Conservative 0; Mismatches 221; Indels 17; Gaps 2:

a 1 ctccaaaggatcccttttggtccttgctcgtcctatcctatggttgcctctctcaaa 60
b 38 ctccaaggatcccttttggtccttgctcgtcctatcctatggttgcctctctcaaa 97

```

[illegible]

GenCore version 4.5
Copyright (C) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 19, 2002, 07:21:40 : Search time 277.57 seconds
(without alignments)
5392.837 Million cell updates/sec

Title: US-09-763-292-1

Perfect score: 1746

Sequence: 1 cttcaaggatctttttg.....acattattattataataaga 1746

Scoring table: IDENTIFY.MSC

Gapop 10.0, Gapext 1.0

Searched: 930621 seqs, 428652619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Watch 0%

Maximum Watch 100%

Listing first 45 summaries

Database : N.Geneseq.1101.*
1: /SID52/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /SID52/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /SID52/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /SID52/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /SID52/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /SID52/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /SID52/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /SID52/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /SID52/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /SID52/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /SID52/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /SID52/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /SID52/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /SID52/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /SID52/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /SID52/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /SID52/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /SID52/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /SID52/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /SID52/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /SID52/gcgdata/geneseq/geneseq/NA2000.DAT.*
22: /SID52/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1739.6	99.6	1746	19 AAV44153	Human cytochrome P
2	1739.6	99.6	1748	22 AAD12241	Human cytochrome P
3	1738	99.5	1746	17 AAT03708	Cytochrome P450 2C
4	1632.4	93.5	1669	19 AAV12759	Wild type cytochrome
5	1534.6	87.9	1845	20 AAV119925	Mammalian cytochrome
6	1534.6	87.9	1845	20 AAV119925	Cytochrome P450 2C
7	1534.6	87.9	1845	20 AAV119925	Human cytochrome P
8	1529.8	87.6	1854	17 AAT11378	Cytochrome P450 2C
9	1529.8	87.6	1854	17 AAT11378	Human cytochrome P
10	1458.6	83.5	1473	16 AAG87728	Human auxillary cy
11	1458.6	83.5	1473	17 AAT28394	Human cytochrome P

12	1455	83.3	1818	11 AAO03599	Human liver cytoch
13	1435.4	82.3	1892	19 AAV44153	Human cytochrome P
14	1415.2	81.1	8475	22 AAF76964	Genetic construct
15	1413.6	81.0	1447	19 AAV12758	Cytochrome P450 2C
16	1401.8	80.9	1438	22 AAF76956	Human derived cyto
17	1373.6	78.7	1591	20 AAX22716	Human cytochrome P
18	1373.6	78.7	1591	20 AAX22716	Human cytochrome P
19	1373.6	78.7	2829	21 AAZ63333	Partial sequence
20	1343.4	76.9	1473	16 AAO87715	Human cytochrome P
21	1343.4	76.9	1473	16 AAT28381	Human cytochrome P
22	1343.4	76.9	1473	16 AAT17404	Human cytochrome P
23	1276	73.1	1419	20 AAX22716	Human derived cyto
24	1264.6	72.4	2009	19 AAV44155	Cytochrome P450 2C
25	1264.6	72.4	2258	19 AAT11380	Human cytochrome P
26	1264.6	72.4	2258	19 AAV44158	Cytochrome P450 2C
27	1254.6	72.4	1473	16 AAO87726	Human cytochrome P
28	1254.6	72.4	1473	16 AAO87726	Human cytochrome P
29	1171.8	67.1	1473	16 AAT28391	Human cytochrome P
30	1172.4	67.1	1922	22 AAS7473	Human cytochrome P
31	1170	67.0	1829	17 AAT11375	Human cytochrome P
32	1170	67.0	1829	19 AAV44156	Human cytochrome P
33	1109.8	63.6	1473	16 AAO87724	Human auxillary cy
34	1109.8	63.6	1473	17 AAT28390	Human cytochrome P
35	1109.8	63.6	1473	17 AAT17394	Human derived cyto
36	1105.6	63.4	1473	16 AAO87725	Human auxillary cy
37	1105.6	63.4	1473	17 AAT28391	Human cytochrome P
38	1105	63.3	1473	16 AAO87726	Human auxillary cy
39	1105	63.3	1473	17 AAT28391	Human cytochrome P
40	645	31.3	1356	21 AAT19392	Human cytochrome P
41	631	36.1	1950	21 AAT38122	Bacterial and mann
42	627.4	35.9	1479	22 AAF76958	Human cytochrome P
43	627.2	35.9	1482	16 AAO87716	Human derived cyto
44	627.2	35.9	1482	17 AAT28382	Human cytochrome P
45	627.2	35.9	1482	17 AAT17414	Human derived cyto

ALIGNMENTS

RESULT 1
AAV44153
ID AAV44153 standard; cDNA: 1746 BP.
XX
AC AAV44153:
XX
DI 06-OCT-1998 (first entry)
XX
DE Human cytochrome P450 2C19 clone 11a cDNA.
XX
XX Cytochrome P450: drug screening; S-mephenytoin 4'-hydroxylase;
XX Identical; identical; identical; identical; identical; identical;
XX Identical; identical; identical; identical; identical; identical;
XX xenobiotic; environmental pollutant; ss.
XX
XX Homo sapiens.
XX
XX US5786191-A.
XX
XX 28-JUL-1998.
XX
XX 22-FEB-1994: 94US-0201118.
XX
XX 22-FEB-1994: 94US-0201118.
XX
XX 09-APR-1992: 92US-0864962.
XX
XX (GOLD/) GOLDSTEIN J A.
XX (ROMK/) ROMKES-SPARKS M.
XX
XX Goldstein JA, Romkes-sparks M;
XX WPI: 1998-436528/37.
XX
XX Screening for drugs metabolised by cytochrome P450 - for identifying
XX mutagenic, carcinogenic, or cytotoxic compounds
XX

CC In the presence of guanosine cofactor. The present sequence is human
CC cytochrome P450 (CYP450) 2C19 cDNA related to the invention.
XX
XX Sequence 1748 BP: 478 A; 419 C; 365 G; 486 T; 0 other.

	Query Match	99.6%	Score 1739.56	DB 22;	Length 1748:
	Best Loc.	Similarity	99.81%	Pred. No. 0:	
	Matches 1742:	Conservative		H mismatches 4:	Indels 0: Gaps
OY	1	c t c t c a a g a g a t c t t t g t g t c t c t g t g t g t c t c a t g t c t c a t g t c t c t c t c t c t c a a	50		
Dδ	3	c t c t c a a g a g a t c t t t g t g t c t c t g t g t c t c a t g t c t c a t g t g t c t c t c t c t c a a	52		
OY	61	t c t g a g a c a g a g c t c t g g a g a g a a a c t c c t c t g t g c c c a c t c t c t c c a g t a	120		
Dδ	63	t c t g a g a c a g a g c t c t g g a g a g a a a c t c c t c t g t g c c c a c t c t c t c c a g t a	122		
OY	121	t t g p n a a t a t c a c g a t a g n a t t a o g g a t f c a g a c a a t c e l t a a c c a a t c t c t c a a	180		
Dδ	123	t t g p n a a t a t c a c g a t a g n a t t a o g g a t f c a g a c a a t c e l t a a c c a a t c t c t c a a	182		
OY	181	a a a t a t c c c c t g t t c a c t c t g a t t t g t g c c t g a c a g t g a g t g a g t c a t a g	240		
Dδ	183	a a a t a t g g c c t g t t c a c t c t g a t t t t g g c t g g a c a g a t g t g t g t c t g a t	242		
OY	241	g a t a g a a g t g t g a g a g a g c c t g a t t g a t c t g a a g a g a g t t t c t g a a g a g c c	300		
Dδ	243	g a t a g a a g t g t g a g a g a g c c t g a t t g a t c t g a a g a g a g t t t c t g a a g a g c c	302		
OY	301	a t t t c c c a t g g t g a a g a g c t a a c a g g a t t g g a a t c g t t t c a g a c a a t g g a a g a	350		
Dδ	303	a t t t c c a t g g t g a a g a g c t a a c a g g a t t g g a a t c g t t t c a g a c a a t g g a a g a	352		
OY	361	g a t g a g a g a g a t c g g g t t t c c t c a t a g e g c t g c g a a t t t g a g a t g g g a a g a	420		
Dδ	363	g a t g a g a g a g a t c g g g t t t c c t c a t a g e g c t g c g a a t t t t g g a t g g g g a a g a	422		
OY	421	g g a g a t a a g a g a c c t t c a a g a g a a g c c g c t c t t g g a g a g t t g g a g a a a a	480		
Dδ	423	g g a g a t a a g a g a c c t t c a a g a g a a g c c g t c t t g g a g a g t t g a g a a a a a	482		
OY	481	c a a a g a t t a a c c c t g a t c a c c t t c a c c t t g g g t c t c c t c a c a c a t g a t c t	540		
Dδ	483	c a a g g t t c a c c t g a t c c a c t t t c a t c t g g g t g t g t c t c t g a a t g a t c a t	542		
OY	541	g t c c c a a t t t t c c a a a c g t t t g a t a t a a a g a c a g a a t t c t t a a c t a a t g g	600		
Dδ	543	g t c c c a a t t t t c c a a a c g t t t g a t a t a a a g a c a g a a t t c t t a a c t a a t g g	602		
OY	601	a a a a t t g a a t g a a a c a t c a g a t t g a a g a c c c t g a t c a g a t a c a a t t	650		
Dδ	603	a a a a t t g a a t g a a a c a t c a g a t t g a a g a c c c t g a t c a g a t a c a a t t	652		
OY	661	t t c c c a c t a t c a t g a t t t c c g g a a c c a a c a a c a a t t a c t t a a a c c t g t t	720		
Dδ	663	t t c c c a c t a t c a t g a t t t c c c g g a a c c a a c a a a t t a c t t a a a a c c t g t t	722		
OY	721	t t a t g a a g t g a t a t t t g g a a a g t a a a a g a c c a a c a a g a t c g a t g a c a t c a a	780		
Dδ	723	t t a t g a a g t g a t a t t t g g a a a g t a a a a g a c c a a c a a g a t c g a t g a c a t c a a	782		
OY	781	a c c t c t g g a c t t a t g a t t c t c t g a c a a a t g g a a a g a a g a a c c a a c c a a c	840		
Dδ	783	a c c t c t g g a c t t a t g a t t c t c t g a c a a a t g g a a a g a a g a a c c a a c c a a c	842		
OY	841	a g t c g a t c c a t a t a a a c t g a a c t g a a c a g a a g a a g a a t t a c t t g a a g a t g a	900		
Dδ	843	a g t c g a t c c a t a t a a a c t g g a t a c a t g a g t g a c t t a c t t g a g t g g a	902		
OY	901	g a g a g a c a a g a c a c c t g a a t g a t c t c t c t c t g t g a a c c a c c a g a g a	950		
Dδ	903	g a g a g a c a a g a c a c c t g a a t g a t c t c t c t c t g t g a a c c a c c a g a g a	952		


```
||||| 486 ccaaggctccctctggtatcccaacttcatctgggtgtgtctccctgcagtgatct 545
Oy 541 gctccatattttccagaagagctttcgattataaagatcacgaattcttaacttgatg 600
Db 546 gctccatattttccagaagagctttcgattataaagatcacgaattcttaacttgatg 605
Oy 601 aaaaatggaagaaacacacgaaggttgaagccctcgatccgattatcagaataat 660
Db 606 aaaaatggaagaaacacacgaaggttgaagccctcgatccgattatcagaataat 665
Oy 661 ttcccaactcatgattatttcccggaacccataaacttacttaaaacacttgctt 720
Db 666 ttctctcatattgatttacttcccggaacccataaacttacttaaaacacttgctt 725
Oy 721 ttatggaagtgatattttggaagataaaagacacacacgaatcgatgcacatcaaca 780
Db 726 ttatggaagtgatattttggaagataaaagacacacacgaatcgatgcacatcaaca 785
Oy 781 accctcggaactttatgatttctctctctctctctctctctctctctctctctctct 840
Db 786 accctcggaactttatgatttctctctctctctctctctctctctctctctctctct 845
Oy 841 agtctgaatttctctctctctctctctctctctctctctctctctctctctctctct 900
Db 846 catctgaatttctctctctctctctctctctctctctctctctctctctctctctct 905
Oy 901 cagagacacacgaacacacacacacacacacacacacacacacacacacacacacac 960
Db 906 cagagacacacgaacacacacacacacacacacacacacacacacacacacacacac 965
Oy 961 tcaacagtaagtcacagagagagagagagagagagagagagagagagagagagagag 1020
Db 966 tcaacagtaagtcacagagagagagagagagagagagagagagagagagagagagag 1025
Oy 1021 tgcacagagggggcacaatccctacacagatgtgtgtgcacagagtcacagagataca 1080
Db 1026 agaatctcccaacacacacacacacacacacacacacacacacacacacacacac 1085
Oy 1081 tgcacagagggggcacaatccctacacagatgtgtgtgcacagagtcacagagataca 1140
Db 1086 tgcacagagggggcacaatccctacacagatgtgtgtgcacagagtcacagagataca 1145
Oy 1141 acctcatctcccaagggcacaacacataataacttctctctctctctctctctctctct 1200
Db 1146 atctcatctcccaagggcacaacacataataacttctctctctctctctctctctctct 1205
Oy 1201 aagaatttcccaacacacacacacacacacacacacacacacacacacacacacac 1260
Db 1206 agaatctcccaacacacacacacacacacacacacacacacacacacacacacac 1265
Oy 1261 tcaagaagaagtaactactctatctctctctctctctctctctctctctctctctctct 1320
Db 1266 tcaagaagaagtaactactctatctctctctctctctctctctctctctctctctctct 1325
Oy 1321 gctgcgcgcgtaggagctgttttattctctgaccttctttacagaactttaacctga 1380
Db 1326 cctgcgcgcgtaggagctgttttattctctgaccttctttacagaactttaacctga 1385
Oy 1381 aatctctgattgacccaagagaccttgacacacacacacacacacacacacacacacac 1440
Db 1386 aatctctgattgacccaagagaccttgacacacacacacacacacacacacacacacac 1445
Oy 1441 tccgcacctctctacagctgtctctctctctctctctctctctctctctctctctctct 1500
Db 1446 tccgcacctctctacagctgtctctctctctctctctctctctctctctctctctctct 1505
Oy 1501 ctctgtgtgtctctctgacgtctctctctctctctctctctctctctctctctctctct 1553
Db 1506 ctgtgtgtgtctctgacgtctctctctctctctctctctctctctctctctctctctct 1555
Oy 1554 tgaatctctctgacccgtctctctctctctctctctctctctctctctctctctctctct 1613
```


XX Cytochrome P450; drug screening; S-mephenytoin 4'-hydroxylase;
KW identification; mutagenic; carcinogenic; cytotoxic; hemoprotein;
XX xenobiotic; environmental pollutant; ss.
XX Homo sapiens.
OS
XX US5786191-A.
XX 28-JUL-1998.
XX 22-FEB-1994; 94US-0201118.
XX 22-FEB-1994; 94US-0201118.
XX 09-APR-1992; 92US-0864962.
XX (GOLD)/ GOLDSTEIN J A.
XX (ROMK)/ ROMKES-SPARKS M.
XX Goldstein JA, Romkes-sparks M;
XX WPI; 1998-436528/37.
XX Screening for drugs metabolised by cytochrome P450 - for identifying
XX mutagenic, carcinogenic, or cytotoxic compounds
XX Example 2; Column 53-56; 63pp; English.
XX This sequence encodes a human cytochrome P450 2C9 polypeptide isolated
XX from clone 65. This polypeptide is used in a method to screen for a drug
XX that is metabolised by a specific cytochrome P450. The protein can also be used to identify a
XX mutagenic, carcinogenic or cytotoxic compound. Cytochrome P450 are a
XX large family of hemoprotein enzymes capable of metabolising xenobiotics
XX such as drugs carcinogens and environmental pollutants as well as
XX endobiotics such as steroids, fatty acids and prostaglandins.
XX Sequence 1852 BP; 514 A; 424 C; 380 G; 534 T; 0 other;
XX
Query Match 87.9%; Score 1534.6; DB 19; Length 1852;
Best Local Similarity 92.9%; Pred. No. 0;
Matches 1637; Conservative 0; Mismatches 109; Indels 17; Gaps 2;
OY 1 cttcaatggtcttttggtgcttctgtctgtctctcatgtttgtctcttctca 60
DB 6 cttcaatggtcttttggtgcttctgtctgtctctcatgtttgtctcttca 65
OY 61 tctggagacagctctggagaggaagaaactccctctgcccactctctccca 120
DB 66 tctggagacagctctggagaggaagaaactccctctgcccactctctccca 125
OY 121 ttggaataatctcacaatagatataagagatgcagcaaatcttaacaaatctctca 180
DB 126 ttggaataatctcacaatagatataagagatgcagcaaatcttaacaaatctctca 185
OY 181 aaatctatgctctgttcaactgtattttggctggagcgaatggtgctgcatg 240
DB 186 aggtctatgctctgttcaactgtattttggctggagcgaatggtgctgcatg 245
OY 241 gatataagatggtggaagcctctgattatctcttgagagagagattttctggaagagcc 300
DB 246 gatataagatggtggaagcctctgattatctcttgagagagagattttctggaagagcc 305
OY 301 atttccactggtcgaagagcgaacagagatttgaatcgttttcagcaaggaaga 360
DB 306 ttttccactggtcgaagagcgaacagagatttgaatcgttttcagcaaggaaga 365
OY 361 gatggaagagattcggcgtttctctccatgacgctgaggaatttggaatggaggaaga 420
DB 366 aatggagagagattcggcgtttctctccatgacgctgaggaatttggaatggaggaaga 425
OY 421 ggaagcattgagaccctgttcaagaggaagccgcgtgctctgtggagaggttgaga 480

DB 426 ggaagcattgagaccctgttcaagaggaagccgcgtgctctgtggagaggttgaga 485
OY 481 ccaaggtcttcaactggtgaccccttctctctggtgctgctccctgcaatgact 540
DB 486 ccaaggtcttcaactggtgaccccttctctctggtgctgctccctgcaatgact 545
OY 541 gttccatttatttccagaacacttctgattataaagatcaagaatttctaactgag 600
DB 546 gttccatttatttccagaacacttctgattataaagatcaagaatttctaactgag 605
OY 601 aaaaattgaatgaacaacacagagatttgaagcaccctctggtatccagatgaataatt 660
DB 606 aaaaattgaatgaacaacacagagatttgaagcaccctctggtatccagatgaataatt 665
OY 661 tttccactatattgattatttccgggaaccattatacaaaatttacttaaaacttgctt 720
DB 666 tttccactatattgattatttccgggaaccattatacaaaatttacttaaaacttgctt 725
OY 721 ttatgagaagatatttttggagaactataaagaacacacacagaatcagtgacatca 780
DB 726 ttatgagaagatatttttggagaactataaagaacacacacagaatcagtgacatca 785
OY 781 accctcgagactttattgatttctctctgataaaattggagaaggaagaaacacacac 840
DB 786 accctcgagactttattgatttctctctgataaaattggagaaggaagaaacacacac 845
OY 841 agctcgaattcttcaatgaataacttgatgaactcagctcagttgacttlttgagagctgga 900
DB 846 catcgaatttacttataaagcttggaaacactcagttgacttlttgagagctgga 905
OY 901 cagagacaagaacacacacacttgagatgctctctctctctgctgaagcaccagagg 960
DB 906 cagagacaagaacacacacacttgagatgctctctctctctgctgaagcaccagagg 965
OY 961 tccgcgttaagctccagagagagattgagcgtgctcatttgcagaagaacacagccctcgc 1020
DB 966 tccgcgttaagctccagagagagattgagcgtgctcatttgcagaagaacacagccctcgc 1025
OY 1021 tgcacagagggccacatgccctcacagatgctgtgtgcacgaggtccagagataca 1080
DB 1026 tgcacagagggccacatgccctcacagatgctgtgtgcacgaggtccagagataca 1085
OY 1081 tgaactctatcccacacagctgctcccaatgcagtgacctgtgacgttaaatccagaact 1140
DB 1086 tgaactctatcccacacagctgctcccaatgcagtgacctgtgacgttaaatccagaact 1145
OY 1141 acctcattcccaagagcacaacatataaacttctctcaetctctgctacatgacaaca 1200
DB 1146 acctcattcccaagagcacaacatataaacttctctgacttctgctacatgacaaca 1205
OY 1201 aagaatttcccaacacagagatgttgaacctgctcaactttctgagatgaagtggaatt 1260
DB 1206 aagaatttcccaacacagagatgttgaacctgctcaactttctgagatgaagtggaatt 1265
OY 1261 ttaagaagaatgaactctgacgtcctctctcagcaggaagaaacagatltgtggagagg 1320
DB 1266 ttaagaagaatgaactctgacgtcctctctcagcaggaagaaacagatltgtggagagg 1325
OY 1321 gcttgcgcgcagagagctgttttattctgaccttcaatttacaagaacttcaacctga 1380
DB 1326 ccttgcgcgcagagagctgttttattctgaccttcaatttacaagaacttcaacctga 1385
OY 1381 aatctctgattgaacccaagagccttgacaaactcctctgtgtcaatgagatttgcctcg 1440
DB 1386 aatctctgattgaacccaagagccttgacaaactcctctgtgtcaatgagatttgcctcg 1445
OY 1441 tcccgccttctatcagctgtgcttcaattctctgtcagagaacacagatggtgctg 1500
DB 1446 tpcgccttctatcagcgtgtgttctctctgctcagagaagcagatggctgctg 1505
OY 1501 cttctgtgctctcctcagctccttctctctctg-----tccaaatttcaatctg 1553


```

Db 788 accctcaggacttatgattgctctctcctgataaaatggaggaagaaagacacacac 847
Oy 841 agtctgaattcctattgaaacttggaactcactcagctgacttactcttgagctggga 900
De 848 cactcgaattactcattgaagcttggaacacactcagctgacttcttgagctggga 907
Oy 901 caagagcaacacacacacacacacacacacacacacacacacacacacacacacac 960
De 908 caagagcaacacacacacacacacacacacacacacacacacacacacacacacac 967
Oy 961 tccagctaaagtcaggaagagatgaaacgtgtcattgagcaaaacacacacacacac 1020
Db 968 tccagctaaagtcaggaagagatgaaacgtgtcattgagcaaaacacacacacacac 1027
Oy 1021 tgcagcagaggggcccacatgccctcaccagatgctgtgtgacagaggtccagagataca 1080
De 1028 tgcagcagaggggcccacatgccctcaccagatgctgtgtgacagaggtccagagataca 1087
Oy 1081 tgcagctcattccacacacacacacacacacacacacacacacacacacacacacac 1140
De 1088 tgcagctcattccacacacacacacacacacacacacacacacacacacacacacac 1147
Oy 1141 acctcattccacacacacacacacacacacacacacacacacacacacacacacacac 1200
De 1148 atctcattccacacacacacacacacacacacacacacacacacacacacacacacac 1207
Oy 1201 agaatctccacacacacacacacacacacacacacacacacacacacacacacacacac 1250
De 1208 agaatctccacacacacacacacacacacacacacacacacacacacacacacacacac 1257
Oy 1261 taaagaaagttaactcattcattcattcattcattcattcattcattcattcattcattc 1320
De 1268 taaagaaagttaactcattcattcattcattcattcattcattcattcattcattcattc 1327
Oy 1321 gctcggccgagtcagctgctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1380
De 1328 cctcggccgagtcagctgctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1387
Oy 1381 aactcctgagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtc 1440
De 1388 aactcctgagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtc 1447
Oy 1441 tccgcctctctcattcagctgctcattcattcattcattcattcattcattcattcattc 1500
De 1448 tccgcctctctcattcagctgctcattcattcattcattcattcattcattcattcattc 1507
Oy 1501 ctctcgtgctgctcgcagctcctctctctctctctctctctctctctctctctctctct 1553
De 1508 ctgctgagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtc 1557
Oy 1554 tgaatcctctcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtc 1613
De 1561 tgaatcctctcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtc 1627
Oy 1614 cagctcctcata-----aaagttctcagtcagtcagtcagtcagtcagtcagtcagtc 1663
De 1628 cagctcctcatacagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtc 1687
Oy 1664 tactctaatagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtc 1723
De 1688 tactctaatagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtc 1747
Oy 1724 taaacattattataaataaga 1746
De 1748 taaatattattataaataaga 1770

RESULT 9
ANV4154
Db ANV4154 standard; cDNA: 1854 BP.
Oy
De
ANV4154:

```

```
Db 368 aatggaagagatccggcgttctccctcctcagcgtcggaatttggagtgaggaga 427
Oy 421 gagacattgagacccgtgtccaagagagcccgctgcttggaggggttaagaasaa 480
Db 428 gagacattgagacccgtgtccaagagagcccgctgcttggaggggttaagaasaa 487
Oy 481 ccaagcgttccacccgtgtccacccgtgtccacccgtgtccacccgtgtccacccgt 540
Db 488 ccaagcgttccacccgtgtccacccgtgtccacccgtgtccacccgtgtccacccgt 547
Oy 541 gttccattttccagaacacgttgcattataaagatcacagcaatttcttaacttgag 600
Db 548 gttccattttccagaacacgttgcattataaagatcacagcaatttcttaacttgag 607
Oy 601 aanaattgaatgaacacagcattgtaagcaccctctggtacccagatgacaaatt 660
Db 608 aanaattgaatgaacacagcattgtaagcaccctctggtacccagatgacaaatt 667
Oy 661 tccacattcattgatttcccggaagacccatcaacaatttcaaaacttgcctt 720
Db 668 tttctctcattgatttcccggaagacccatcaacaatttcaaaacttgcctt 727
Oy 721 tatggaagtgatatttggagaagttaaaagacacccaagaatcgatgacatcaaca 780
Db 728 tatggaagtgatatttggagaagttaaaagacacccaagaatcgatgacatcaaca 787
Oy 781 accctcgagcatttattgattgtctctgataaaaatggagaggaagcaaaacacac 840
Db 788 accctcgagcatttattgattgtctctgataaaaatggagaggaagcaaaacacac 847
Oy 841 agctgaattcatttgaacattgtgatacctcagctcagctcagctcagctcagctgga 900
Db 848 catctgaatttacttgaagcttggaacacacacacacacacacacacacacacacac 907
Oy 901 cagagacacacacacacacacacacacacacacacacacacacacacacacacacacac 960
Db 908 cagagacacacacacacacacacacacacacacacacacacacacacacacacacacac 967
Oy 961 tcaagcattcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1020
Db 968 tcaagcattcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1027
Oy 1021 tgcacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1080
Db 1028 tgcacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1087
Oy 1081 tgcacacacacacacacacacacacacacacacacacacacacacacacacacacacac 1140
Db 1088 tgcacacacacacacacacacacacacacacacacacacacacacacacacacacacac 1147
Oy 1141 accctcattcccaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1200
Db 1148 attcattcccaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1207
Oy 1201 aagaatttcccaacccagagatgttgacccctgctcactttcttgagtgaggtggaatt 1260
Db 1208 aagaatttcccaacccagagatgttgacccctgctcactttcttgagtgaggtggaatt 1267
Oy 1261 taagaagtgatcacttctcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1320
Db 1268 taagaagtgatcacttctcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1327
Oy 1321 gcttgccgcatgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1380
Db 1328 ccttgccgcatgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1387
Oy 1381 aatctcattgaccccaagacacacacacacacacacacacacacacacacacacacacac 1440
Db 1388 aatctcattgaccccaagacacacacacacacacacacacacacacacacacacacacac 1447
Oy 1441 tcccgcttctcactcagctgctcactcctgctgctgaggaagcagcagcagcagcagcagc 1500
```



```

PH Key: Location/Qualifiers
FT CD5 1..1443
TT /*tag= a
XX JP02072879-A.
XX 13-MAR-1990.
XX 09-SEP-1988: 88JP-0225955.
XX 09-SEP-1988: 88JP-0225955.
XX (AGEN ) AGENCY OF IND SCI TECH.
XX WPI: 1990-121045/16.
XX P-PSDB; AAR04043.
XX Recombinant plasmid, for yeast for liver treatment -
XX contains human liver cytochrome P-450MP gene obtd. from yeast.
XX etc.
XX PS
XX Disclosure: Fig 4: 9pp; Japanese.
XX
XX The P-450 produced from plasmid pHp6 in a yeast expression system
XX preferably Saccharomyces AH22, can be used to treat human liver
XX disorders, oxidising various chemical substances.
XX
XX Sequence 1818 BP: 516 A; 412 C; 376 G; 514 T; 0 other;
XX
Query Match 83.3%; Score 1455; DB 11; Length 1818;
Best Local Similarity 92.3%; Pred. No. 0;
Matches 1574; Conservative 0; Mismatches 110; Indels 23; Gaps 3:
OY 57 tcaatctggagcagcgtctgtggaggaagaaactctctctggccccactctctccca 116
DB 31 tgaatctggagcagcgtctgtggaggaagaaactctctctggccccactctctccca 90
OY 117 gtgattgaaatctctcagatagatattaaagatgacagcaaatctctcaacaaatc 176
DB 91 gtgattgaaatctctcagatagatattaaagatgacagcaaatctctcaacaaatc 150
OY 177 tcaaaatctatggccgtgtgtctcctctgtatttggcttggacacatgtgtgtgtg 236
DB 151 tcaaaatctatggccgtgtgtctcctctgtatttggcttggacacacatgtgtgtg 210
OY 237 catgatgatgaagtgtgaaggaagccctgattgattctggaggaagatttctgtgga 296
DB 211 catgatgatgaagcagcaggaagccctgattgattctggaggaagatttctgtgga 270
OY 297 ggcatttccacatgctgctgctgctgctgctgctgctgctgctgctgctgctg 356
DB 271 ggcatttccacatgctgctgctgctgctgctgctgctgctgctgctgctgctg 330
OY 357 aagaatgagaaggaatgctgctgctgctgctgctgctgctgctgctgctgctgctg 416
DB 331 aagaatgagaaggaatgctgctgctgctgctgctgctgctgctgctgctgctgctg 390
OY 417 aagagagcattgagcagcgtgttcaagaggaagccgctgctgtggagaggttgaga 476
DB 391 aagagagcattgagcagcgtgttcaagaggaagccgctgctgtggagaggttgaga 450
OY 477 aaaaacaaagctcaacctgtgctgctgctgctgctgctgctgctgctgctgctg 536
DB 451 aaaaacaaagctcaacctgtgctgctgctgctgctgctgctgctgctgctgctg 510
OY 537 atctgtccatatttccagaacgcttctgattataaagatcagcaatttcttaacttg 596
DB 511 atctgtccatatttccagaacgcttctgattataaagatcagcaatttcttaacttg 570
OY 597 atggaaaattgaataacatcaggaggtgaagcaccacccctggatccagatagcaat 656
DB 571 atggaaaattgaataacatcaggaggtgaagcaccacccctggatccagatagcaat 630

```

```

OY 657 aatttccactatcattatatttccgggaacccataaataattcttaaaacctt 716
DB 631 aatttctctactatcattatatttccgggaacccataaataattcttaaaacctt 690
OY 717 gctttatggaagtgatattttgggaaggtataaagaacacacacacacacacac 776
DB 691 gctttatggaagtgatattttgggaaggtataaagaacacacacacacacacac 750
OY 777 acaacccctcgagacttattgattctctctgatacaaatgagagaagaaacacac 836
DB 751 acaacccctcgagacttattgattctctctgatacaaatgagagaagaaacacac 804
OY 837 caacagctgaattcactatttgaataacacagctgacgtgacgttacttggagct 896
DB 805 caacagctgaattcactatttgaataacacagctgacgtgacgttacttggagct 864
OY 897 ggaacaggaacaaacacacacacacacacacacacacacacacacacacacac 956
DB 865 ggaacaggaacaaacacacacacacacacacacacacacacacacacacacac 924
OY 957 gagctcagctaaagctcagagagagattgaacgtgattgagcagaacccggagccc 1016
DB 925 gagctcagctaaagctcagagagagattgaacgtgattgagcagaacccggagccc 984
OY 1017 tgaatcagcagacacacacacacacacacacacacacacacacacacacacac 1076
DB 985 tgaatcagcagacacacacacacacacacacacacacacacacacacacacac 1044
OY 1077 taatcagcactcctccacacagctgcccacgtcagtgacgtgacgttataaattcaga 1136
DB 1045 taatcagcactcctccacacagctgcccacgtcagtgacgtgacgttataaattcaga 1104
OY 1137 aactactcactccacagcagacacacacacacacacacacacacacacacacac 1196
DB 1105 aactactcactccacagcagacacacacacacacacacacacacacacacacac 1164
OY 1197 acaaaagatttccacacacacacacacacacacacacacacacacacacacac 1256
DB 1165 acaaaagatttccacacacacacacacacacacacacacacacacacacacac 1224
OY 1257 aattttaagaaagtaactctcactgcttctcagcagggaaaacggattctgtgga 1316
DB 1225 aattttaagaaagtaactctcactgcttctcagcagggaaaacggattctgtgga 1284
OY 1317 gagggcctgcccagcagtgagctgttttctcctgaccttctcatttccagaacttaac 1376
DB 1285 gagggcctgcccagcagtgagctgttttctcctgaccttctcatttccagaacttaac 1344
OY 1377 ctgaatctctgattgaccccaagacccctcacaacacactcctgtgtcgaatgtgct 1436
DB 1345 ctgaatctctgattgaccccaagacccctcacaacacactcctgtgtcgaatgtgct 1404
OY 1437 tctatcccccctctcactgacgtctcctcctcctcctcctcctcctcctcctc 1496
DB 1405 tctatcccccctctcactgacgtctcctcctcctcctcctcctcctcctcctc 1464
OY 1497 gctgctcctgtgctcctcctcctcctcctcctcctcctcctcctcctcctcctc 1549
DB 1465 gctgctcctgtgctcctcctcctcctcctcctcctcctcctcctcctcctcctc 1524
OY 1550 tctgtaagcttctctgacccgacacacacacacacacacacacacacacacacac 1609
DB 1525 tctgtaagcttctctgacccgacacacacacacacacacacacacacacacacac 1584
OY 1610 catcagcctccacta-----aaaaagtttcaactgtgcaaatatattctgtattc 1659
DB 1585 catcagcctccacta-----aaaaagtttcaactgtgcaaatatattctgtattc 1644
OY 1660 cccactactataatagtttcaatgagctcacaataaagctgacttctgtatgttga 1719
DB 1645 tccaactctgacagctgctgactgactgactgactgactgactgactgactgactg 1704

```


Db 848 caacagcaacacccctgagatgctctctctctctctgctgagagccacccagaggtccag 907
Oy 967 ctaagtcacaggaagagatgagctgctcattgctgagaaacccgagccctcgaagtcacg 1026
Db 908 ctaagtcacaggaagagatgagctgctcattgctgagaaacccgagccctcgaagtcacg 967
Oy 1027 aaggggcaacatgcccctacacagatgctgctgagagaggtccagagatcaatcagacc 1086
Db 968 cagggggcacaatgcccctacacagatgctgctgagagaggtccagagatcaatcagacc 1027
Oy 1087 tcatcccacacagctgcccctacacagatgctgctgagagaggtccagagatcaatcagacc 1146
Db 1028 tcatcccacacagctgcccctacacagatgctgctgagagaggtccagagatcaatcagacc 1087
Oy 1147 tcccacagggcaacacatataacttccctcactctctgctgctacatgacacacaaagaa 1206
Db 1088 tcccacagggcaacacatataacttccctcactctctgctgctacatgacacacaaagaa 1147
Oy 1207 tcccacagggcaacacatataacttccctcactctctgctgctacatgacacacaaagaa 1265
Db 1148 tcccacagggcaacacatataacttccctcactctctgctgctacatgacacacaaagaa 1207
Oy 1267 aaggaactactctgctctctcctcagcaggaacacagatgctgctgagagaggtccg 1326
Db 1208 aaggaactactctgctctctcctcagcaggaacacagatgctgctgagagaggtccg 1267
Oy 1327 cccgcatgagctgtttattctcctgactctcattttacacagaactttacacgaacttc 1385
Db 1268 cccgcatgagctgtttattctcctgactctcattttacacagaactttacacgaacttc 1327
Oy 1387 tgaatgacccaaagacactgacacacactctgtgtcaatgattgtgtgtgtgtgtgtgt 1446
Db 1328 tgaatgacccaaagacactgacacacactctgtgtcaatgattgtgtgtgtgtgtgtgt 1387
Oy 1447 cctctctcagctgtctctcattctcctgctgctgagagacacagatgctgtgtgtgtgtgt 1505
Db 1388 cctctctcagctgtctctcattctcctgctgctgagagacacagatgctgtgtgtgtgtgt 1447
Oy 1507 tgcct 1510
Db 1448 gctc 1451

RESULT 15
AAV12758
ID AAV12758 standard; cDNA; 1447 BP.
XX AC AAV12758;
XX UT 08-MAY-1998 (first entry)
XX DE Cytochrome P450 2C19 mutin gene.
KW Cytochrome P450 2C19; human; uninterrupted metabolism; omeprazole;
KW diazepam; imipramine; miltin; ss.
XX OS Homo sapiens.
XX FH Location/Qualifiers
XX CDS 4..1419
XX /tag a
XX /transl_except- (pos: 382..384, aa:Asn)
XX /transl_except- (pos: 495..498, aa:Asn)
XX /transl_except- (pos: 520..522, aa:Asn)
XX /transl_except- (pos: 523..525, aa:Asn)
XX /transl_except- (pos: 585..588, aa:Asn)
XX /transl_except- (pos: 700..702, aa:Asn)
XX /transl_except- (pos: 772..774, aa:Asn)
XX /transl_except- (pos: 778..780, aa:Asn)
XX /transl_except- (pos: 781..783, aa:Asn)
XX /transl_except- (pos: 916..918, aa:Asn)
XX /transl_except- (pos: 964..966, aa:Asn)
XX /transl_except- (pos: 1012..1014, aa:Asn)

Query Match 81.0%; Score 1413.6; DB 19; Length 1447;
Best Local Similarity 99.0%; Pred. No. 0;
Matches 142; Conservative 0; Mismatches 14; Indels 0; Gaps 0;
Oy 67 gacagagctcggagagagaaactcctctgcccactctctccacagatggaa 126
Db 11 gacatctcggagagagaaactcctctgcccactctctccacagatggaa 70
Oy 127 atactctacagatattagagatgacacaaactcttaaccaactctcaaaatct 186
Db 71 atactctacagatattagagatgacacaaactcttaaccaactctcaaaatct 130
Oy 187 atggcctctgtctcactctgtattttgagcgaacgcatggtgctgcatgatatg 246
Db 131 atggcctctgtctcactctgtattttgagcgaacgcatggtgctgcatgatatg 190
Oy 247 aatggtgaagagagccctgattgattcttgagagagagattttctgagagagccattcc 306
Db 191 aatggtgaagagagccctgattgattcttgagagagagattttctgagagagccattcc 250
Oy 307 cactggtgaagagagcctacacagagatttgaatcgattttcagcaatgagaaagatgga 366
Db 251 cactggtgaagagagcctacacagagatttgaatcgattttcagcaatgagaaagatgga 310
Oy 367 aggaagattcgcctctctcctcctcagcctgaggaattttggatggggaagaggagca 426
Db 311 aggaagattcgcctctctcctcctcagcctgaggaattttggatggggaagaggagca 370
Oy 427 ttagagagccgttctcagagagagccctcctctgtgagagagatttgaagaaacccaagg 486
Db 371 ttagagagccgttctcagagagagccctcctctgtgagagagatttgaagaaacccaagg 430
Oy 487 cttaacccctgagtcctccttctcctgagcgtgctcctcgaatgctgctctca 546
Db 431 cttaacccctgagtcctccttctcctgagcgtgctcctcgaatgctgctctca 490
Oy 547 ttatttccagaaacgtttcagattataagatcacgaatttcttaactgagtggaat 606
Db 491 ttatttccagaaacgtttcagattataagatcacgaatttcttaactgagtggaat 550
Oy 607 tgaatgaaacatcagagattgaagccctcctgagcagatgacatatttttccca 666
Db 550 tgaatgaaacatcagagattgaagccctcctgagcagatgacatatttttccca 614

Claim 4: Page 9-10; 15pp: Japanese.
XX This sequence represents the modified version of the human liver derived
XX cytochrome P450 2C19 gene of the invention. The modifications comprise
XX a substitution of the second amino acid with Ala and a deletion
XX of amino acid residues 3-20 of the wild type. Modified cytochrome
XX P450 2C19 is useful for: uninterrupted metabolism of therapeutics,
XX e.g. omeprazole, diazepam and imipramine in humans.
XX
XX Sequence 1447 BP; 410 A; 341 C; 322 G; 374 T; 0 other:

Db 551 tgaatgaatacaatcagattgtaagcaccctcctggtacagagatgcaataattttccca 610
Oy 557 ctatcattgattatttcccggaacccatacaaaattacttaaaaccttctgttttaagg 726
Db 511 ctatcattgattatttcccggaacccatacaaaattacttaaaaccttctgttttaagg 670
Oy 727 aaagtgaatttttgggaanaagaaacacccaaagatcgatggacatcaacaacccctc 786
Db 571 aaagtgaatttttgggaanaagaaacacccaaagatcgatggacatcaacaacccctc 730
Oy 787 ggaactttatgatttctctctgatacaaaatggaagaagaaacacacacacactctg 846
Db 731 ggaactttatgatttctctctgatacaaaatggaagaagaaacacacacacactctg 790
Oy 847 aattcactatgaaaacttgaatcactcagctgacttacttggagctggagacagaga 906
Db 791 aattcactatgaaaacttgaatcactcagctgacttacttggagctggagacagaga 850
Oy 907 caacaagaacacacactgagataagctctctctctctctctctctctctctctctct 966
Db 851 caacaagaacacacactgagataagctctctctctctctctctctctctctctctct 910
Oy 967 ctaaagtccagagagagatgaaactgctcatttggcagaacccgagccctgcatgcag 1026
Db 911 ctaaagtccagagagagatgaaactgctcatttggcagaacccgagccctgcatgcag 970
Oy 1027 acaggggccaatgcctacacagatgctgtgtgacagaggctccagagatacatcgacc 1084
Db 971 acaggggccaatgcctacacagatgctgtgtgacagaggctccagagatacatcgacc 1030
Oy 1087 tcataccaccacagctcccatgacgtgactgtgacgttaaatccagaactaccctca 1146
Db 1031 tcataccaccacagctcccatgacgtgactgtgacgttaaatccagaactaccctca 1090
Oy 1147 ttcccaaggggcaacacacataataacttccctcacttctgtgctacatgacacaaagaat 1206
Db 1091 ttcccaaggggcaacacacataataacttccctcacttctgtgctacatgacacaaagaat 1150
Oy 1207 ttcccaaccacagagagattgaactcctgactcacttctgtgctacatgacacaaaga 1266
Db 1151 ttcccaaccacagagagattgaactcctgactcacttctgtgctacatgacacaaaga 1210
Oy 1267 aaagttaactacttctcagcttctcagcaggaacacagatttgtgtggagagggccctg 1326
Db 1211 aaagttaactacttctcagcttctcagcaggaacacagatttgtgtggagagggccctg 1270
Oy 1327 ccgcatgagctgtttttatctctgacttcaatttcaagaactttaactgaaatctc 1386
Db 1271 ccgcatgagctgtttttatctctgacttcaatttcaagaactttaactgaaatctc 1330
Oy 1387 tgattgaccacaaagacacttgacacacactcctgtgtcaatgatttctgtctccgc 1446
Db 1331 tgattgaccacaaagacacttgacacacactcctgtgtcaatgatttctgtctccgc 1390
Oy 1447 ccttctatcagctgtcttctcattctctgactgagagacagatggtctgagctgct 1502
Db 1391 ccttctatcagctgtcttctcattctctgactgagagacagatggtctgagagct 1446

Search completed: April 19, 2002, 08:15:21
Job time: 3221 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 19, 2002, 07:21:40 ; Search time 2723.34 Seconds
4889.382 Million cell updates/sec

Title: US-09-763-292-1
Perfect score: 1746
Sequence: 1 cttcaatgctctttgtg.....acattattattataaga 1746

Scoring table: IDENTITY_MUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 537288281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum Match 100%

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database :

EST.*

1: em_estfun.*
2: em_esthum.*
3: em_estin.*
4: em_estom.*
5: em_estpl.*
6: em_estba.*
7: em_estro.*
8: em_estov.*
9: em_hic.*
10: em_hic2.*
11: gb_est2.*
12: gb_hic.*
13: gb_gss.*
14: em_gss.fun.*
15: em_gss.hum.*
16: em_gss.inv.*
17: em_gss.pin.*
18: em_gss.pro.*
19: em_gss.vit.*
20: em_gss.vit.*
21: em_gss.other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed.
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	984.8	56.4	1990	12 AK008580	AK008580 Mus muscu
2	966.6	55.4	1825	12 AK005017	AK005017 Mus muscu
3	965	55.3	1870	12 AK008688	AK008688 Mus muscu
4	788	45.1	1775	12 AK008512	AK008512 Mus muscu
5	699.2	40.0	909	11 BG198899	BG198899 RST18150
6	672.6	38.5	802	11 BG217233	BG217233 RST18150
7	664	37.5	779	11 BG215664	BG215664 RST18150
8	654	37.5	779	11 BG215664	BG215664 RST18150
9	626	35.9	893	11 BG195327	BG195327 RST18150
10	619	35.5	791	11 BG198890	BG198890 RST18150
11	606.4	34.7	1017	10 A1529366	A1529366 u167e10.y
12	592.4	33.9	910	11 BG198348	BG198348 RST17505

13	589.8	33.8	907	11 BG205076	BG205076 RST24495
14	587.4	33.6	856	11 BG186520	BG186520 RST5487.A
15	585.2	33.5	667	10 AV653206	AV653206 AV653206
16	584.8	33.5	777	11 BG186022	BG186022 RST1978.A
17	578.8	33.2	899	11 BG202484	BG202484 RST21842
18	569	32.5	613	10 AV655841	AV655841 AV655841
19	564	32.5	876	11 BG196822	BG196822 RST16049
20	552.8	31.5	712	10 AV659165	AV659165 AV659165
21	552.8	31.5	712	10 AV659165	AV659165 AV659165
22	550.2	31.5	782	10 AV653830	AV653830 AV653830
23	547	31.3	901	11 BG205598	BG205598 RST25061
24	546.4	31.3	620	10 AV647944	AV647944 AV647944
25	543.4	31.1	864	11 BG211904	BG211904 RST31479
26	537.2	30.8	973	11 BG566700	BG566700 602585761
27	530.8	30.4	928	11 BG205597	BG205597 RST25060
28	530	30.4	632	10 AV654640	AV654640 AV654640
29	530	30.4	678	10 AV651762	AV651762 AV651762
30	527.6	30.2	726	11 BG211595	BG211595 RST18150
31	527.6	30.2	726	11 BG211595	BG211595 RST18150
32	514	29.4	829	10 A1047349	A1047349 u165c04.y
33	511.4	29.3	824	10 A1255553	A1255553 u155h09.y
34	511.2	29.3	821	10 A1047293	A1047293 u164e07.y
35	510.4	29.2	830	10 A1097740	A1097740 u163c07.y
36	507.8	29.1	787	10 A1181879	A1181879 u163c01.y
37	506.8	29.0	716	10 AV648351	AV648351 AV648351
38	503	28.8	867	11 BG195966	BG195966 RST15049
39	495.6	28.6	863	11 BG128308	BG128308 602985620
40	495.6	28.6	559	10 AV647947	AV647947 AV647947
41	487.8	28.3	720	10 A1097528	A1097528 u165c01.y
42	456.6	28.2	720	10 A1097528	A1097528 u165c01.y
43	456.2	28.4	934	11 BG200448	BG200448 RST19761
44	455.2	28.4	841	10 A1097729	A1097729 u163b05.y
45	494	28.3	813	11 BG185988	BG185988 RST1943.A

ALIGNMENTS

RESULT	AK008580	1990 bp	HTC	05-JUL-2001
LOCUS	Mus musculus adult male small intestine cDNA, RIKEN full-length enriched library, clone:2010318C06, full insert sequence.			
DEFINITION	AK008580	1990 bp	HTC	05-JUL-2001
ACCESSION	AK008580	1990 bp	HTC	05-JUL-2001
VERSION	AK008580.1	GI:12842852		
KEYWORDS	CAP trapper.			
SOURCE	Mus musculus (strain:G57H/63) adult male small intestine cDNA to mRNA, clone:2010318C06, RIKEN full-length enriched mouse cDNA library			
ORGANISM	Mus musculus			
REFERENCE	Eukaryote: Metazoa: Chordata: Vertebrata: Euteleostomi: Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Murinae: Mus.			
AUTHORS	Carninci, P. and Hayashizaki, Y.			
TITLE	High-efficiency full-length cDNA cloning			
JOURNAL	Methods in enzymology. 303, 19-44 (1999)			
REFERENCE	1 (bases 1 to 1990)			
AUTHORS	Itou, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.			
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes			
JOURNAL	Genome research. 10 (10), 1617-1630 (2000)			
MEDLINE	20499374			
PUBMED	11042159			
REFERENCE	2 (bases 1 to 1990)			
AUTHORS	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itou, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.			
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes			
JOURNAL	Genome research. 10 (10), 1617-1630 (2000)			
MEDLINE	20499374			
PUBMED	11042159			
REFERENCE	3 (bases 1 to 1990)			
AUTHORS	Shikama, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitzunai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J.,			

Okazaki, Y., Muramatsu, M., Inoue, Y., Kiro, A. and Hayashizaki, Y. RIKEN integrated sequence analysis (RISA) system—384-format sequencing pipeline with 384 multipipillary sequencer. *Genome Research*, 10 (11), 1757-1771 (2000). 20530913

11076861

The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium. Functional annotation of a full-length mouse cDNA collection. *Nature* 409, 685-690 (2001). 5 (bases 1 to 1825)

Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Aono, H., Arai, A., Arakawa, T., Carninci, P., Fukuda, S., Fukunishi, Y., Goto, M., Horii, F., Hwang, T., Imai, A., Itoh, M., Iwano, M., Izawa, M., Kato, H., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Shibata, K., Shibata, Y., Shingawa, A., Shiraki, T., Sogabe, T., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T., Tezuka, I., Toyota, T., Yamamura, T., Yamashita, A., Yoshida, K., Yoshino, M., Aizumitsu, M. and Hayashizaki, T. Submitted (10-June-2000) Yoshihide Hayashizaki. The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@sc.riken.go.jp). URL: <http://genome.gsc.riken.go.jp/>, Tel: 81-43-503-3222, Fax: 81-43-503-3223. Web site: <http://genome.gsc.riken.go.jp/> for

please visit our web site (<http://genome.gsc.riken.go.jp/>) for further details.

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Center in Riken. The library was screened with a probe derived from *Usp1* cDNA contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5'] GAGAGAGACGGCGCCGCACTGCAGTCGATTTTITTTT 3', cDNA was prepared by using tobacco thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer [3'] GAGAGAGACGGCGCCGCACTGCAGTCGATTTTITTTT 5' and the primer [5'] GAGAGAGACGGCGCCGCACTGCAGTCGATTTTITTTT 3'. cDNA was cloned with XhoI and SalI. Cloning sites, 5' end: SalI; 3' end: XhoI. Host: λ gt10.

```

1. 1825
Location/Qualifiers
/organism="Mus musculus"
/strain="C57BL/6J"
/sex="male"
/db_xref="MGI:13061009"
/db_xref="MGI:13061806"
/db_xref="MGI:1904427"
/clone="1300015K14"
/sex="male"
/tissue_type="liver"
/tissue_type="spleen"
/tissue_type="full-length enriched mouse cDNA library"

```

```

18. 1490      /note="putative"
      /codon_start=1
      /protein_id="BAB3748.1"
      /translation="MDPIVLVLTISCLLSLAWROSSERKLPGETPLPIGNILQ
      /db_xref="gi:12836645"
      IDKATNIGCFNLSVAGVGVPIVTLGKPTVLEKMGKRSIEDVKEEAFAGRGRTV
      FVDAKINQKQVAFKSNQKVVNTRFSLKRLNGKMGKRSIEDVKEEAFAGRGRTV
      GSPCDPTFLGCGAQCWIGSIIIFDDFVDKMDIKMLKRLNITFLKSSVQKQDQ
      TPALDDYFLGCGAQCWIGSIIIFDDFVDKMDIKMLKRLNITFLKSSVQKQDQ
      NQPLNQLKHLAIVLIDVLSAGTETSIIIPNSIVPTKMLKRVKQVEEIEVHVGK
      SPQCDKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQ
      KQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQ
      SLQWFLKFLKPLVHPDIDVTPMLGLIASVPPAQLCFIPS"
1804. 1809
      /note="putative"
      /codon_start=1
      /protein_id="BAB3748.1"
      /translation="MDPIVLVLTISCLLSLAWROSSERKLPGETPLPIGNILQ
      /db_xref="gi:12836645"
      IDKATNIGCFNLSVAGVGVPIVTLGKPTVLEKMGKRSIEDVKEEAFAGRGRTV
      FVDAKINQKQVAFKSNQKVVNTRFSLKRLNGKMGKRSIEDVKEEAFAGRGRTV
      GSPCDPTFLGCGAQCWIGSIIIFDDFVDKMDIKMLKRLNITFLKSSVQKQDQ
      TPALDDYFLGCGAQCWIGSIIIFDDFVDKMDIKMLKRLNITFLKSSVQKQDQ
      NQPLNQLKHLAIVLIDVLSAGTETSIIIPNSIVPTKMLKRVKQVEEIEVHVGK
      SPQCDKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQ
      KQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQ
      SLQWFLKFLKPLVHPDIDVTPMLGLIASVPPAQLCFIPS"
1804. 1809

```

```

_signal
1804. 1809
/note='putative'

```



```

|||||
908 AGTCAACCCAGGTCACAGAGATGACGTGTGTGATTCGACAGACAGAGGCCCTG 967
Oy 1019 catcacagagagccacatccctacacatgctggtggtgacagagtcagagata 1078
|||||
968 CATCGAGGAGGACATGATGCCCTACACTGATGCTGTGTCGATGATGATGATGAT 1027
Oy 1079 catgacctatccccacagcctgcccacatgcagctgacctgacgttaaatcaaaa 1138
|||||
1080 CATGATCTCTCCACAGCCTGCCATGCGGTGACCTGACCTGACCTGACCTGACCTG 1087
Oy 1139 catcctatctccacagcagcagcattatcctcctcctcctcctcctcctcctc 1198
|||||
1088 GTATCTCATTCACAGGACCATCTGATAGCATGACTGACTGCTGTGCTGATATGATGA 1147
Oy 1199 caaagaatttcccaaccacagagatgttgacctgctcactcttctgataagagtgaa 1258
|||||
1148 TAAAGAAATTCACACACAGAGATTTGACCCAGTCTGATCTGATGATGATGATGAT 1207
Oy 1259 tttaagaagaagtaacttcaatctcctcctcctcctcctcctcctcctcctcct 1318
|||||
1208 GTTAAAGAAATTCACACACAGAGATTTGACCCAGTCTGATGATGATGATGATGAT 1267
Oy 1319 gggctgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1378
|||||
1268 GGGCTGTGCTCGACGAGCAATATTTTCTGCTGACCATCTGATGATGATGATGAT 1377
Oy 1379 gaattctctgattgacccaagagcctgacacacactctgtgtcaatggattgtcttc 1438
|||||
1378 GAATCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1387
Oy 1439 ggtcagcctctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 1479
|||||
1388 TGTGCCACCAAGTCCAGATATGCTTCATCCCACTMAA 1428

RESULT 4
LOCUS AK008512 1775 bp mRNA HTC 05-JUL-2001
DEFINITION Mus musculus adult male small intestine cDNA, RIKEN full-length
ACCESSION AK008512.1 GI:12842736
VERSION 1
KEYWORDS CAP trapper.
SOURCE Mus musculus (strain: C57BL/6J) adult male small intestine cDNA to
mRNA, clone lib=RIKEN full-length enriched mouse cDNA library
clone:201030IM18.
ORGANISM Mus musculus
Mikoyota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus;
MUS.
REFERENCE 1 (bases 1 to 1775)
AUTHORS Carninci, P. and Hayashizaki, Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Methods in enzymology. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE 2 (bases 1 to 1775)
AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Kanno, H., Kiyama, J., Nishi, K., Katsuno, M., Hayashizaki, Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome research. 10 (10), 1617-1630 (2000)
MEDLINE 20493374
PUBMED 11042159
REFERENCE 3 (bases 1 to 1775)
AUTHORS Shibata, K., Itoh, M., Aizawa, K., Nagata, S., Sasaki, N., Carninci, P.,
Kanno, H., Kiyama, J., Nishi, K., Katsuno, M., Hayashizaki, Y.,
Kumai, M., Shimada, K., Kato, H., Itoh, M., Hayashizaki, Y.,
Fujisake, S., Inoue, K., Togawa, M., Izawa, M., Ohara, E., Katsuki, M.,
Yoneda, Y., Ishikawa, T., Tanaka, T., Tanaka, T., Matsura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format

```

```

sequencing pipeline with 384, multicapillary sequencer
Genome research. 10 (11), 1757-1771 (2000)
20530933
11076861
4 (bases 1 to 1775)
The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 585-590 (2001)
Adachi, A., Akao, Y., Akahira, S., Akimura, T., Aono, H., Arai, A.,
Arakawa, T., Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M.,
Hanagaki, T., Hara, A., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hori, F.,
Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kato, H., Kawai, J.,
Kojima, Y., Kondo, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T.,
Miyazaki, A., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y.,
Okido, T., Owa, C., Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H.,
Sasaki, D., Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T.,
Sogabe, Y., Suzuki, H., Tabata, M., Tagawa, A., Takahashi, F.,
Tadaka, Y., Tejima, F., Toyama, T., Yamamura, K., Yamashita, A.,
Yasuda, K., Yoshida, K., Yuzawa, M., Aizawa, M. and Hayashizaki, Y.
Direct Submission
Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp,
URL: http://genome-gsc.riken.go.jp/, Tel: 81-45-503-9222.
Fax: 81-45-503-9216)
Please refer our web site (http://genome-gsc.riken.go.jp/) for
further details.
FEATURES
Location/Qualifiers
1..1775
Source
/organism="Mus musculus"
/db_xref="taxon:10090"
/db_xref="MIM:1902075"
/db_xref="MIM:1917138"
/size="1775"
/sex="male"
/tissue_type="small intestine"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
BASE COUNT 509 a 382 c 343 g 541 t
ORIGIN
Query Match 45 11: Score 788; DB 12: Length 1775;
Best Local Similarity 80.74; Pred. NO. 2.9e-204;
Matches 920; Conservative 0; Mismatches 220; Indels 0; Gaps 0;
Oy 340 tcgtttcagcaatggaagagatgagagatcggtttctcctcctcctcctcctcctcctc 399
|||||
Db 194 TGTGTCATGAATGAGAAACCAAGAGTGTACCTGTGATCCCACTTCATTCCTGGCT 253
|||||
Oy 400 ggaatttggatgggaagagagatgagagcgtttcagaagagagagagagagagagag 459
|||||
Db 254 GAATTTTGGATGGGAGAGAGATGATGAGGACCGATTCAGAGGAGGCTCGCTGTC 313
|||||
Oy 460 ttgtggagagtggaagaaacacagcctcctcctcctcctcctcctcctcctcctcct 519
|||||
Db 314 TGTGTCATGAATGAGAAACCAAGAGTGTACCTGTGATCCCACTTCATTCCTGGCT 373
|||||
Oy 520 gtcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 579
|||||
Db 374 GTCCTCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTC 433
|||||
Oy 580 acgaatttcttaacttgatggaagaaatgaaagaaacatcaggatgaaagcctcctc 639
|||||
Db 434 AGGAATTTCTTACCTTTATAGACATCTTAAATGAAATTTAGAAATTTCTGAGTTCCCT 493
|||||
Oy 640 ggaacagatgcaaaatttcccaatcattgattatttccggggaacacataca 699
|||||

```


Db 494 GGATTCAGGTTTGCATTAATTTCCCGCTCATTTGATTATCTACCGAGGACAGAA 553
Oy 700 aatLactataaaactgtgtttatggaagtgatatttggaagaagtaaaagcaacc 759
Db 554 AATTTACTTAGAAATTTTGATATTGCAAAACATTTACTTTTGGCAAGTATACAGATA 613
Oy 760 agaaatgataggtgatcaaaacctcgagcttattgatgtctctgatacaatgg 819
Db 614 AAGAGTACTGGATATATACATCTCGGACATTTTGTATTTTCTGTGATCAAAATGG 673
Oy 820 agaggaagaaagcaaaacacagctgtgaattcactattgaacacttgaaactcagc 879
Db 674 AGCAGGAAAGCAACCAACCAAGAGAGGATTTACTTGTGAGAACTTGATCTCAGCAT 733
Oy 880 ctgacttacttgaggtggagagagacaaacacacctcgagatgtgctctcttc 939
Db 734 CTGATCTGTGTCTGTGGAGAGACAGACAGCAACCCCTGAGTACTCTGTCTGTC 793
Oy 940 tctgtactgaagccagagctgcagactaaactcagagagaagatgaacgtgctg 999
Db 914 TGATGAGATCAGAGATACATGATCTCTTCCACACAGCTTCCCATGCGCTGACCC 973
Oy 1120 gtagcgttaattcgaacactcactcattcccaaggagcaacacatttaacttctc 1179
Db 974 GTGAGTGAATTCAGAGAGTATCTCATTTCCCAAGGCAACCACTGTATAGCACTGA 1033
Oy 1180 ctctgtgctacatgacaaagaattctcccaacccagagatgtttgacctgtcaact 1239
Db 1034 CTCTGTGCTATATGATGATTAAGAAATCTCAACCCAGAGAGTGTGATCTAGTACT 1093
Oy 1240 ctctgtgaggtggaagatattgaagaagaaactcactcactcgtctctcagagaa 1299
Db 1094 TTCTGAGTGAAGTGGCAGCTTTTGAAGAAAGTGACTACTTTTTTCTTCTCAACAGAA 1353
Oy 1300 aacgagttgtgtgagagaggtgctgcccagcagagctgttttattctgacctca 1359
Db 1154 AAGGATTTGTGTGGAGAGTGTCTGGCTCGAACAGAAATTTATTTTATCTTGACACGA 1213
Oy 1360 ttctcgaacttbaactgaattctcgtatgacccaaagagaccttgacacactctg 1419
Db 1214 TTTTACAGAACTTTACTTGAATCTCAGTTTGTATCTGAAGACCTTGATCTACTCCAG 1273
Oy 1420 ttgcaatgatttctgtcgtccgctctctatcagctgtgcttctctctgctga 1479
Db 1274 TTGCAATGGATTTGTTCTGTGTCACCAAGTTCAGATATGCTTCTTCTCACTAAA 1333

RESULT 5
LOCUS R5718160
DEFINITION R5718160 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
ACCESSION BG198889
VERSION BG198889.1 GI:13720576
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE (bases 1 to 30)
AUTHORS Rain,S., Dahl,T., Thornton,M., Ramachandran,B., Whittington,J., Lerner,L., Krashinsky,D., McElligott,K., Clark,S., Mays,R., Smith,E., Veloso,N., Hess,J., Cothren,K., Lo,K., Offenbacher,J., Danzig,J., and Ducar,M.
TITLE Creation of Genome-wide Protein Expression Libraries using Random

Activation of Gene Expression
Nat. Biotechnol. 19 (5): 440 (2001) in press
Contact: Scott J. Cain
Athersys, Inc.
3201 Carnegie Ave., Cleveland, OH 44115, USA
Tel: 216 431 9500
Fax: 216 361 8225
Email: scott.cain@atersys.com
High quality sequence stop: 532.
Location/Qualifiers
source
1. 909
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Athersys RAGE Library"
/cell_line="H1080"
/note="See 'Creation of Genome-wide Protein Expression Libraries using Random Activation of Gene Expression' in press. Note that even though the cell type indicated is H1080, since a random activation method was used, these sequence tags are not necessarily expressed in H1080 under normal circumstances."
BASE COUNT 237 a 240 c 181 g 250 t
ORIGIN
Query Match 40.0%; Score 659.2; DB J1: Length 909;
Best Local Similarity 88.5%; Fields No. 6e-180;
Matches 787; Conservative 0; Mismatches 84; Indels 18; Gaps 2;
Oy 822 aagaaagagaaacacacagctctgaattcactattgaacacttgtaatacctcagct 881
Db 5 AGCAAAAGCCCAACCACTCTGTAATTTACTATTGAAGCTTGGAAACACTGCAGT 64
Oy 882 gacttacttgaggtggagagagacaaacacacctcgagatgtgctctctctc 941
Db 65 GACTCTGTGGAGTGGAGAGAGACAGACAGCAACCCCTGAGATATCTCTCTCTC 124
Oy 942 ctactgaacacacacagctcagactcaactaaagtcagagagagatgaagctgctatgc 1001
Db 125 CTGCTGAAGCAGGAGAGGTCACAGTAAAGTCCAGAGAGATTGAACTGTGATGGC 184
Oy 1002 gaaacacggagccctctgatacagacagggggccacatgcccacagatgctggg 1061
Db 185 AGAAACCGAGAGCCCTGCATGAGACAGAGGCCATGCCCTACACAGATGCTGGTG 244
Oy 1062 cagagctcagagatcactcagactcactccacagctgacctgacctgaactgt 1121
Db 245 CACGAGCTCCAGAGATACATTGACTCTCCACAGAGCTGGCCCTGCACTGCTGT 304
Oy 1122 gacgttaattcagaactcactcattcccaaggagcaacacatatttaacttccctact 1181
Db 305 GACATTAATTCAGAACTATCTCATTCACAGGCAACACATATTAATTTCCCTGACT 364
Oy 1182 tctgtacatgacaaagaagattcccaacccagagatgttgacctctcaacttt 1241
Db 365 TCTGTGATGACAGCAAGAGATTTCCCAACCAAGATGTTTGACCTCATCACTTT 424
Oy 1242 ctgagtaggtgggaatttttaagaagaagtaagtaactctatgactcttctcagcagga 1301
Db 425 CTGATGAAGGTGGCAATTTTAAGAAAGTAATAACTTCAATGCTTTCTCAGCAGAAA 484
Oy 1302 cagatttggtagggagggcctggccagcagtgagctgtttttattctgaccttcaatt 1361
Db 485 CGAATTTGTGTGGAGAGAGCCCTGGCCGCTAGAGCTGTTTTTATCTTGACCTCCATT 544
Oy 1362 ttacagactttacactgaactctctgattgacctcaacaaagacctgacacactcctgt 1421
Db 545 ATACAGACTTTTACCTGAATCTCTGTTGTGATCCCAAGAACCTTGACACCACTCCAGTT 604
Oy 1422 g:caagtaggttctctgtcccgcccttctatcagctgtgcttctcctctgctgaaga 1481
Db 605 GTCAATGGATTTGCTGTGGCGGCTCTTACAGCTGTGCTTTCATTCCTGCTGTGAAGA 664

[illegible]

	LOCUS	BC186520	856 bp	mRNA	EST	21-APR.-2001			
	DEFINITION	BSTR7549 Athysarsy Rage Library Homo sapiens CDNA mRNA sequence.							
	ACCESSION	BC186520							
	VERSION	BC186520.1	GI:13708207						
	KEYWORDS	EST.							
	SOURCE	human.							
	ORGANISM	Homo sapiens							
	REFERENCE	Eukaryotes; Metazoa; Chordata; Craniata; Vertebrates; Euteleostomi; Mammalia; Primates; Catarrhini; Hominoidea; Hominidae; Homo. 1 (PubMed ID 856)							
	AUTHORS	Barrington,J.J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R., Cain,S., Dahl,T.T., Thornton,M., Ramachandran.R., Whittington,J.E., Lerner,L., Krashinsky,D., McElligott,K., Clark,S., Mays,R., Saith,E., Veloso,H., Resas,J., Cothren,K., Lo.K., Offenbacher,J., Dandajr,J., and Ducar.M.							
	TITLE	Creation of Genome-wide Protein Expression Libraries using Random Activation of Gene Expression							
	JOURNAL	Nucleic Acids Research							
	COMMENT	Volume 29(15), 440 (2001) In press Contact: Scott J. Cain Athysarsy, Inc.							
		3201 Carnegie Ave, Cleveland, OH 44115, USA Tel.: 216 431 9900 Fax: 216 361 9596 Email: scaint@athysarsy.com High quality sequence clonop: 410.							
	FEATURES	Location/Qualifiers ..+ 856 /organism.."Homo sapiens" /db_xref.."taxon:9606" /clone_ilb.."Athysarsy RAGE Library" /cell_line.."HT1080" /note.."See "Creation of Genome-wide Protein Expression Libraries Using Random Activation of Gene Expression" the Nature Biotechnology issue HT1080 since a random activation method was used, these sequence tags are not necessarily expressed in HT1080 under normal circumstances."							
	BASE COUNT	216 a 229 c 175 g 235 t							
	ORIGIN								
		Query Match	33.6%	Score 587.4	PB 11:	Length 856;			
		Best Local Similarity	85.0%;	Penalty -7e-9;					
		Matches 694;	Conservative	0	Mismatches 112;	Indels 10;	Gaps 3;		
OY	861	aacttggaatcaactgcagctgaccttacttggtcggtagacagcacacaaggccaacc	920						
Dd	7	AMGCTGGAAACAAGTGCATGTTCATTGTTTGAGTGTGGACAGACAGCAAGAACC	66						
OY	921	ctcgatatgtcttccttcctctgctaagaacccagaggtcacagttaattgcaggaa	190						
Dd	67	CGAATAATGCTCTTCCTCCCTCGAACGCCAAGGCTTAAGCTTAGACTTCGAAG	286						
OY	981	ggagatgatgcgtcatgccggaacccagagcccctgcatgcacagagggaccatg	1040						
Dd	127	GAGATFGAAGCTGTGATTGGCAAAAACCGAGCCCCTCATGATCAAGACAGAGCACATG	186						
OY	1041	ccttacacagatgctgttgtgcagaggttcagagatacatactgacctctccccaacgc	1100						
Dd	187	CCTCAAACCMATGCTGTGTGCACAGAGTCCAGAGTTCATGTACTCTCCCCACAG	246						
OY	1101	ctgcacctcagctgcagttgcgtttaaatctcaaactaccttatctccccaggacca	1160						
Dd	247	CTGCGCCCATGCAATGCACTTGACATTAATTCAGAAACTATCTCATTCCTCCAAGG	306						
OY	1161	acctataatactctcctactctctgtgctacatgacacaaaagaattctcccaaccagg	1220						
Dd	307	ACCAATATATTCCTGATCTTGTGCTCACTCAAGCAMCAAMAANTATTCCTCCAC	366						
OY	1221	agtgtgaactgcgttcctcttcgtatgtagtgaataatttaaagaagaactactctc	1280						
Dd	367	ATSTTTGACCTTCACTCTCTGSAAGAAGTGCAATTTAGAAAAAGTAAATTAATCTT	1425						

Qy	1281	atgcctllctcagcagaaacacagatctgtggagagagcctgaccccatcgaacgtc	1340
Db	427	ATGCTCTTCTCAGAGAAACCGATTTGTGTGGAGAAAGCCTGCGCCGCATGGAGCTG	486
Qy	1341	ttttttattctcgaactcatttttaacgaacttaacctaagaattctctgattgacccaag	1400
Db	487	TTTTTATTCTCGAGCTCCATTTTACAGAACTTTAACTCTGAAATCTCTGNTGACCCAAG	546
Qy	1401	gaatttgacacaaactctctgttcaatggatttgccttctgtccgcctctctctcagtg	1460
Db	547	AAGCTTTGACACATCCTGCGTGTCAATGGTTTGGCTTGTGTCGCCCTCTTACCAGTG	606
Qy	1461	tacattctctgtctcgttcagagacagcagatgctgcctgcctcagctgcctcag	1520
Db	607	TGCTTCATCTGCTCTGAGAA-AACAGATGGCTGGTGTCTGCTGTGCATCTCCATCGCA	665
Qy	1521	ctctctctctctctg-----tccaaatttcaactatcgtgatgctctctcgcacgt	1573
Db	666	CTCTCTTTGCTTGTGGGGGCTATCACTCTTTAACTATCTGTAAATGC--CTTGCTCACTG	723
Qy	1574	catcacaatttcccttcccccaaatcatagtaaatctcgctcctcctataaaagatt	1633
Db	724	CATCTCAGATTCCTGTCGAGAACTAGCAATTCACCCATTCAGGGAGTTTCTATG	783
Qy	1634	tcaatcagcaaatatcgtctatccccatctt	1669
Db	784	TTATCTGCAATTTCTGTATCTCTACTTTTACACT	819

RESULT	15			
AV553206	557 bp	MRNA	EST	07-SEP-2000
ACCURATION	AV553206	GLC	Homo sapiens	CDNA clone GLCID1E03.3', mRNA sequence.
ACCESSION	AV553205			
VERSION	AV553205.1	GI:9874220		
KEYWORDS	EST.			
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.			
	1 (bases 1 to 557)			
	Qian,B., Wu., Huang,Q., Huang,C., Kang,B., Gao,X., Xu,Z., Xiao,H.			
	Guan,B., Yu., Huang,P., Liu,F., Ou,J., Song,H., Cheng,Z., Ou,J.,			
	Zeng,L., Xu,S., Gu,W., Lu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu,			
	G., Yang,Y., Gu,Y., Chen,Z. and Han,Z.			
TITLE	Homo sapiens CDNA clone			
JOURNAL	Unpublished (2000)			
COMMENT	Contact: Zeguag Han			
	Chinese National Human Genome Center at Shanghai			
	351 Guo Shouling Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai			
	201203, P. R. China			
	Tel: 86-21-50801322			
	Fax: 86-21-50801322 (ex.45)			
	E-mail: hanzgq@ac.sh.cn			
	Excell: hanzgq@ac.sh.cn			
	This clone is available at CHGC in Shanghai.			
FEATURES	Location/Qualifiers			
source	1..657			
	/organism="Homo sapiens"			
	/db_xref="taxon:9606"			
	/clone="GLCID1E03"			
	/clone_lib="GLC"			
	/tissue_type="Corresponding non cancerous liver tissue"			
	/sub_page="01"			
	/note="vector: pBluescript sk(-); Site_1: EcoRI; Site_2: XhoI"			
BASE COUNT	193 a 126 c 151 g 192 t 5 others			
ORIGIN				

Query Match 33.5%; Score 585.2; DB 10; Length 667;
Best Local Similarity 92.1%; Pred. NO. 6.3e-149;
Matches 61A; Conservative 0; Mismatches 53; Indels 0; Gaps 0;


```

OY 159 aaatccttaacaaatctctcaaaaatctatggccctgtgtcactctgtattttggcctg 218
Db 1 aaatccttaacaaatctctcaaaaatctatggccctgtgtcactctgtattttggcctg 60
OY 219 gaacacatggtggtgagatgagatgagatgagatgagatgagatgagatgagatgag 278
Db 61 aaacccatagtggtggtgagatgagatgagatgagatgagatgagatgagatgagatgag 120
OY 279 gaagagttttctggaagagagagagagagagagagagagagagagagagagagagag 318
Db 121 caggagttttctggaagagagagagagagagagagagagagagagagagagagagag 180
OY 339 atcgctttcagcaatggaagagagagagagagagagagagagagagagagagagagag 398
Db 181 atgtgtttcagcaatggaagagagagagagagagagagagagagagagagagagagag 240
OY 399 cgaatgtttgagagagagagagagagagagagagagagagagagagagagagagagag 458
Db 241 cgaatgtttgagagagagagagagagagagagagagagagagagagagagagagagag 300
OY 459 ctgtgagagagagagagagagagagagagagagagagagagagagagagagagagag 518
Db 301 ctgtgagagagagagagagagagagagagagagagagagagagagagagagagagag 360
OY 519 tgtgctcctgcaatgagatgagatgagatgagatgagatgagatgagatgagatgagat 578
Db 361 tgtgctcctgcaatgagatgagatgagatgagatgagatgagatgagatgagatgagat 420
OY 579 cagcaatgttcaactgtgagagagagagagagagagagagagagagagagagagagag 638
Db 421 cagcaatgttcaactgtgagagagagagagagagagagagagagagagagagagagag 480
OY 639 tggatccagatgagatgagatgagatgagatgagatgagatgagatgagatgagatgag 698
Db 481 tggatccagatgagatgagatgagatgagatgagatgagatgagatgagatgagatgag 540
OY 699 aatttacttaaaacctgtctttatggaagagagagagagagagagagagagagagagag 758
Db 541 aatttacttaaaacctgtctttatggaagagagagagagagagagagagagagagagag 600
OY 759 caagagatgagatgagatgagatgagatgagatgagatgagatgagatgagatgagatg 818
Db 601 caagagatgagatgagatgagatgagatgagatgagatgagatgagatgagatgagatg 660
OY 819 gagaagg 825
Db 661 gaaaggg 667

```

Search completed: April 19, 2002, 08:07:44
 Job time: 2764 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 19, 2002, 11:06:25 ; Search time 5830.82 Seconds
(without alignments)
1200.021 Million cell updates/sec

Title: US-09-763-292-3
Perfect score: 323
Sequence: 1 ccccgaaatctacacac.....aaattacatcatctttt 323

Scoring table: IDENTITY NYC

Gapop 10.0, Gapext 1.0

Searched: 22023303 seqs, 10831430700 residues 44046506

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending Patents, NA, Main.*

- 1: /cgn2_6/ptodata/2/pna/US05 COMB.seq.*
- 2: /cgn2_6/ptodata/2/pna/US05 COMB.seq.*
- 3: /cgn2_6/ptodata/2/pna/US07 COMB.seq.*
- 4: /cgn2_6/ptodata/2/pna/US08 COMB.seq.*
- 5: /cgn2_6/ptodata/2/pna/US08 COMB.seq.*
- 6: /cgn2_6/ptodata/2/pna/US08 COMB.seq.*
- 7: /cgn2_6/ptodata/2/pna/US08 COMB.seq.*
- 8: /cgn2_6/ptodata/2/pna/US08 COMB.seq.*
- 9: /cgn2_6/ptodata/2/pna/US08 COMB.seq.*
- 10: /cgn2_6/ptodata/2/pna/US08 COMB.seq.*
- 11: /cgn2_6/ptodata/2/pna/US08 COMB.seq.*
- 12: /cgn2_6/ptodata/2/pna/US08 COMB.seq.*
- 13: /cgn2_6/ptodata/2/pna/US08 COMB.seq.*
- 14: /cgn2_6/ptodata/2/pna/US09 COMB.seq.*
- 15: /cgn2_6/ptodata/2/pna/US09 COMB.seq.*
- 16: /cgn2_6/ptodata/2/pna/US09 COMB.seq.*
- 17: /cgn2_6/ptodata/2/pna/US09 COMB.seq.*
- 18: /cgn2_6/ptodata/2/pna/US09 COMB.seq.*
- 19: /cgn2_6/ptodata/2/pna/US09 COMB.seq.*
- 20: /cgn2_6/ptodata/2/pna/US09 COMB.seq.*
- 21: /cgn2_6/ptodata/2/pna/US09 COMB.seq.*
- 22: /cgn2_6/ptodata/2/pna/US09 COMB.seq.*
- 23: /cgn2_6/ptodata/2/pna/US09 COMB.seq.*
- 24: /cgn2_6/ptodata/2/pna/US09 COMB.seq.*
- 25: /cgn2_6/ptodata/2/pna/US09 COMB.seq.*
- 26: /cgn2_6/ptodata/2/pna/US09 COMB.seq.*
- 27: /cgn2_6/ptodata/2/pna/US09 COMB.seq.*
- 28: /cgn2_6/ptodata/2/pna/US09 COMB.seq.*
- 29: /cgn2_6/ptodata/2/pna/US09 COMB.seq.*
- 30: /cgn2_6/ptodata/2/pna/US09 COMB.seq.*
- 31: /cgn2_6/ptodata/2/pna/US09 COMB.seq.*
- 32: /cgn2_6/ptodata/2/pna/US09 COMB.seq.*
- 33: /cgn2_6/ptodata/2/pna/US09 COMB.seq.*
- 34: /cgn2_6/ptodata/2/pna/US09 COMB.seq.*
- 35: /cgn2_6/ptodata/2/pna/US09 COMB.seq.*
- 36: /cgn2_6/ptodata/2/pna/US09 COMB.seq.*
- 37: /cgn2_6/ptodata/2/pna/US09 COMB.seq.*
- 38: /cgn2_6/ptodata/2/pna/US09 COMB.seq.*
- 39: /cgn2_6/ptodata/2/pna/US09 COMB.seq.*
- 40: /cgn2_6/ptodata/2/pna/US09 COMB.seq.*
- 41: /cgn2_6/ptodata/2/pna/US09 COMB.seq.*
- 42: /cgn2_6/ptodata/2/pna/US09 COMB.seq.*
- 43: /cgn2_6/ptodata/2/pna/US09 COMB.seq.*

- 44: /cgn2_6/ptodata/2/pna/US6011 COMB.seq.*
- 45: /cgn2_6/ptodata/2/pna/US6012 COMB.seq.*
- 46: /cgn2_6/ptodata/2/pna/US6013 COMB.seq.*
- 47: /cgn2_6/ptodata/2/pna/US6014 COMB.seq.*
- 48: /cgn2_6/ptodata/2/pna/US6015 COMB.seq.*
- 49: /cgn2_6/ptodata/2/pna/US6016 COMB.seq.*
- 50: /cgn2_6/ptodata/2/pna/US6017 COMB.seq.*
- 51: /cgn2_6/ptodata/2/pna/US6018 COMB.seq.*
- 52: /cgn2_6/ptodata/2/pna/US6019 COMB.seq.*
- 53: /cgn2_6/ptodata/2/pna/US6020 COMB.seq.*
- 54: /cgn2_6/ptodata/2/pna/US6021 COMB.seq.*
- 55: /cgn2_6/ptodata/2/pna/US6022 COMB.seq.*
- 56: /cgn2_6/ptodata/2/pna/US6023 COMB.seq.*
- 57: /cgn2_6/ptodata/2/pna/US6024 COMB.seq.*
- 58: /cgn2_6/ptodata/2/pna/US6025 COMB.seq.*
- 59: /cgn2_6/ptodata/2/pna/US6026 COMB.seq.*
- 60: /cgn2_6/ptodata/2/pna/US6027 COMB.seq.*
- 61: /cgn2_6/ptodata/2/pna/US6028 COMB.seq.*
- 62: /cgn2_6/ptodata/2/pna/US6029 COMB.seq.*
- 63: /cgn2_6/ptodata/2/pna/US6030 COMB.seq.*
- 64: /cgn2_6/ptodata/2/pna/US6031 COMB.seq.*
- 65: /cgn2_6/ptodata/2/pna/US6032 COMB.seq.*
- 66: /cgn2_6/ptodata/2/pna/US6033 COMB.seq.*
- 67: /cgn2_6/ptodata/2/pna/US6034 COMB.seq.*
- 68: /cgn2_6/ptodata/2/pna/US6035 COMB.seq.*
- 69: /cgn2_6/ptodata/2/pna/US6036 COMB.seq.*
- 70: /cgn2_6/ptodata/2/pna/US6037 COMB.seq.*
- 71: /cgn2_6/ptodata/2/pna/US6038 COMB.seq.*
- 72: /cgn2_6/ptodata/2/pna/US6039 COMB.seq.*
- 73: /cgn2_6/ptodata/2/pna/US6040 COMB.seq.*
- 74: /cgn2_6/ptodata/2/pna/US6041 COMB.seq.*
- 75: /cgn2_6/ptodata/2/pna/US6042 COMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	323	100.0	323	30	US-09-763-292-3
2	323	100.0	323	55	US-60-226-176-759
3	323	100.0	323	56	US-60-233-468-759
4	323	100.0	323	64	US-60-313-371-759
5	310.4	96.1	572	50	US-60-170-373-1355
6	310.4	96.1	572	51	US-60-181-428-187
7	310.4	96.1	572	51	US-60-181-428-187
8	310.4	96.1	572	51	US-60-181-428-187
9	310.4	96.1	572	51	US-60-181-428-187
10	299.4	92.7	10097	54	US-60-213-795-55
11	299.4	92.7	10097	54	US-60-213-795-55
12	299.4	92.7	10097	54	US-60-213-795-55
13	284.2	88.0	3557	53	US-60-207-211-20
14	284.2	88.0	3557	53	US-60-207-211-20
15	284.2	88.0	3557	53	US-60-207-211-20
16	284.2	88.0	3557	53	US-60-207-211-20
17	276.8	85.7	8437	18	US-09-488-127-59
18	276.8	85.7	8437	18	US-09-488-127-59
19	276.8	85.7	8437	18	US-09-488-127-59
20	276.8	85.7	8437	18	US-09-488-127-59
21	275.2	85.2	685	52	US-60-194-110-15
22	275.2	85.2	685	52	US-60-194-110-15
23	275.2	85.2	685	52	US-60-194-110-15
24	275.2	85.2	685	52	US-60-194-110-15
25	275.2	85.2	685	52	US-60-194-110-15
26	275.2	85.2	685	52	US-60-194-110-15
27	275.2	85.2	685	52	US-60-194-110-15
28	263	81.4	348	55	US-60-248-498-25
29	263	81.4	348	55	US-60-248-498-25
30	263	81.4	348	55	US-60-248-498-25
31	263	81.4	348	55	US-60-248-498-25

c 32 263 81.4 56476 57 US-60-245-227-19
c 33 263 81.4 56476 58 US-60-258-278-70
c 34 263 81.4 56476 59 US-60-248-572-73
c 35 254.2 78.8 319 51 US-60-182-895-140
c 36 254.2 78.8 319 51 US-60-182-895-140
c 37 250.4 77.5 495 49 US-60-162-247-1064
c 38 223.4 69.2 13925 57 US-60-248-498-60
c 39 223.4 69.2 13925 57 US-60-248-542-61
c 40 223.4 69.2 15443 54 US-60-213-795-28
c 41 223.4 69.2 36191 57 US-60-248-498-23
c 42 223.4 69.2 36584 57 US-60-245-227-10
c 43 223 69.0 1001 26 US-09-671-317-50
c 44 208.8 64.6 567 52 US-60-198-818-181
c 45 208.8 64.6 567 52 US-60-198-818-181

ALIGNMENTS

RESULT 1
US-09-763-292-3
: Sequence 3, Application US/09763292
: BEST LOCAL SIMILARITY 100.0%; Pred. No. 7.4e-90;
: Matches 323; Conservative 0; Mismatches 0; Indels 0; Gaps 0
: GENERAL INFORMATION:
: APPLICANT: Morris, MacDonald
: TITLE OF INVENTION: Single Nucleotide Polymorphisms Associated With ADME Genes
: FILE REFERENCE: GX-0013-1 P
: CURRENT APPLICATION NUMBER: US/60/226-176
: CURRENT FILING DATE: 2000-08-16
: NUMBER OF SEQ ID NOS: 2447
: SOFTWARE: PERL Program
: SEQ ID NO 3
: LENGTH: 323
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE: misc-feature
: NAME/KEY: misc-feature
: OTHER INFORMATION: GB:HUM2C9X05
US-09-763-292-3

Query Match 100.0%; Score 323; DB 30; Length 323;
Best Local Similarity 100.0%; Pred. No. 7.4e-90;
Matches 323; Conservative 0; Mismatches 0; Indels 0; Gaps 0
Oy 1 cccctgaatgctacacaaatgcccattttctctttccatcagtttttacttg 60
Db 1 cccctgaatgctacacaaatgcccattttctctttccatcagtttttacttg 60
Oy 61 tcttatacagtaaaagtcacgaagagattgaacgtgattggcagaaccggagccct 120
Db 61 tcttatacagtaaaagtcacgaagagattgaacgtgattggcagaaccggagccct 120
Oy 121 gcatgacagcagggccacatgcccctacacagatgctgtgacgaagtcagaagt 180
Db 121 gcatgacagcagggccacatgcccctacacagatgctgtgacgaagtcagaagt 180
Oy 181 acctgaaccttctccaccagcctgcccctgacgtgacgtgacattaaattcagaa 240
Db 181 acctgaaccttctccaccagcctgcccctgacgtgacgtgacattaaattcagaa 240
Oy 241 acctatcattcccaaggtaagttgtttctctacacgtgacattcattttcgaagt 300
Db 241 acctatcattcccaaggtaagttgtttctctacacgtgacattcattttcgaagt 300
Oy 301 cccaaattcatagatcatctttt 323
Db 301 cccaaattcatagatcatctttt 323

RESULT 2
US-60-226-176-759
: Sequence 759, Application US/60226176
: GENERAL INFORMATION:
: APPLICANT: Malsen, Gareth
: APPLICANT: Townley, David
: TITLE OF INVENTION: Single Nucleotide Polymorphisms Associated With ADME Genes
: FILE REFERENCE: GX-0013-1 P
: CURRENT APPLICATION NUMBER: US/60/233-468
: CURRENT FILING DATE: 2000-09-18
: NUMBER OF SEQ ID NOS: 2488
: SOFTWARE: PERL Program
: SEQ ID NO 759
: LENGTH: 323
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE: misc-feature
: NAME/KEY: misc-feature
: OTHER INFORMATION: GB:HUM2C9X05
US-60-226-176-759

APPLICANT: Morris, MacDonald
: TITLE OF INVENTION: Single Nucleotide Polymorphisms Associated With ADME Genes
: FILE REFERENCE: GX-0013-1 P
: CURRENT APPLICATION NUMBER: US/60/226-176
: CURRENT FILING DATE: 2000-08-16
: NUMBER OF SEQ ID NOS: 2447
: SOFTWARE: PERL Program
: SEQ ID NO 759
: LENGTH: 323
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE: misc-feature
: NAME/KEY: misc-feature
: OTHER INFORMATION: GB:HUM2C9X05
US-60-226-176-759

Query Match 100.0%; Score 323; DB 55; Length 323;
Best Local Similarity 100.0%; Pred. No. 7.4e-90;
Matches 323; Conservative 0; Mismatches 0; Indels 0; Gaps 0
Oy 1 cccctgaatgctacacaaatgcccattttctctttccatcagtttttacttg 60
Db 1 cccctgaatgctacacaaatgcccattttctctttccatcagtttttacttg 60
Oy 61 tcttatacagtaaaagtcacgaagagattgaacgtgattggcagaaccggagccct 120
Db 61 tcttatacagtaaaagtcacgaagagattgaacgtgattggcagaaccggagccct 120
Oy 121 gcatgacagcagggccacatgcccctacacagatgctgtgacgaagtcagaagt 180
Db 121 gcatgacagcagggccacatgcccctacacagatgctgtgacgaagtcagaagt 180
Oy 181 acctgaaccttctccaccagcctgcccctgacgtgacgtgacattaaattcagaa 240
Db 181 acctgaaccttctccaccagcctgcccctgacgtgacgtgacattaaattcagaa 240
Oy 241 acctatcattcccaaggtaagttgtttctctacacgtgacattcattttcgaagt 300
Db 241 acctatcattcccaaggtaagttgtttctctacacgtgacattcattttcgaagt 300
Oy 301 cccaaattcatagatcatctttt 323
Db 301 cccaaattcatagatcatctttt 323

RESULT 3
US-60-233-468-759
: Sequence 759, Application US/60233468
: GENERAL INFORMATION:
: APPLICANT: Malsen, Gareth
: APPLICANT: Townley, David
: APPLICANT: Valdes, Ana
: TITLE OF INVENTION: Single Nucleotide Polymorphisms Associated With ADME Genes
: FILE REFERENCE: GX-0013-1 P
: CURRENT APPLICATION NUMBER: US/60/233-468
: CURRENT FILING DATE: 2000-09-18
: NUMBER OF SEQ ID NOS: 2488
: SOFTWARE: PERL Program
: SEQ ID NO 759
: LENGTH: 323
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE: misc-feature
: NAME/KEY: misc-feature
: OTHER INFORMATION: GB:HUM2C9X05
US-60-233-468-759

Query Match 100.0%; Score 323; DB 56; Length 323;
Best Local Similarity 100.0%; Pred. No. 7.4e-90;

```
Matches 323; Conservative 0; Mismatches 0; Indels 0; Gaps 2;
Oy 1 cccctgattgctcaacaagatgagcattttctcttccatcagtttttactgtg 60
Db 1 cccctgattgctcaacaagatgagcattttctcttccatcagtttttactgtg 60
Oy 61 ttttatcagctaaagtccagagagattgacgtgtgattgagcaaacggagccct 120
Db 61 ttttatcagctaaagtccagagagattgacgtgtgattgagcaaacggagccct 120
Oy 121 gcatgcaagacagagagccacatgccctcacagatgctgtgtgcagagttccagagat 180
Db 121 gcatgcaagacagagagccacatgccctcacagatgctgtgtgcagagttccagagat 180
Oy 181 accttgaccttcccacacagctgcccctgacgtgacctgtgacattaaattcagaa 240
Db 181 accttgaccttcccacacagctgcccctgacgtgacctgtgacattaaattcagaa 240
Oy 241 accttctctcccaaggaattttttctctacacagctgacctgtgttttccagat 300
Db 241 accttctctcccaaggaattttttctctacacagctgacctgtgttttccagat 300
Oy 301 ccccaattcatagatcattttt 323
Db 301 ccccaattcatagatcattttt 323

RESULT 4
US-60-313-371-759
: Sequence 187, Application US/6013371
: GENERAL INFORMATION:
: APPLICANT: Ring, Ruijun Z.
: APPLICANT: Malsen, Gareth
: APPLICANT: Tomley, David
: APPLICANT: Morris, MacDonald
: TITLE OF INVENTION: Single Nucleotide Polymorphisms Associated With ADME Genes
: FILE REFERENCE: GR-0013-5 P
: CURRENT APPLICATION NUMBER: US/60/313-371
: NUMBER OF SEQ ID NOS: 08-16
: SOFTWARE: PERL Program
: SEQ ID NO 759
: LENGTH: 323
: TYPE: DNA
: ORGANISM: Homo sapiens
: NAME/REF: misc.feature
: OTHER INFORMATION: GB:HM2C9X05
US-60-313-371-759

Query Match 100.0%; Score 323; DB 64; Length 323;
Best Local Similarity 100.0%; Pred. No. 7.4e-90;
Matches 323; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1 cccctgattgctcaacaagatgagcattttctcttccatcagtttttactgtg 60
Db 1 cccctgattgctcaacaagatgagcattttctcttccatcagtttttactgtg 60
Oy 61 ttttatcagctaaagtccagagagattgacgtgtgattgagcaaacggagccct 120
Db 61 ttttatcagctaaagtccagagagattgacgtgtgattgagcaaacggagccct 120
Oy 121 gcatgcaagacagagagccacatgccctcacagatgctgtgtgcagagttccagagat 180
Db 121 gcatgcaagacagagagccacatgccctcacagatgctgtgtgcagagttccagagat 180
Oy 181 accttgaccttcccacacagctgcccctgacgtgacctgtgacattaaattcagaa 240
Db 181 accttgaccttcccacacagctgcccctgacgtgacctgtgacattaaattcagaa 240
Oy 241 accttctctcccaaggaattttttctctacacagctgacctgtgttttccagat 300
Db 241 accttctctcccaaggaattttttctctacacagctgacctgtgttttccagat 300

RESULT 5
US-60-170-373-1355/c
: Sequence 1355, Application US/60170373
: GENERAL INFORMATION:
: APPLICANT: Bonazzi, Vivien
: TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
: TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS AND USES
: FILE REFERENCE: THERDOP
: CURRENT APPLICATION NUMBER: US/60/170-373
: NUMBER OF SEQ ID NOS: 198-12-13
: SOFTWARE: FASTSEQ For Windows Version 4.0
: SEQ ID NO 1355
: LENGTH: 572
: TYPE: DNA
: ORGANISM: Human
US-60-170-373-1355

Query Match 96.1%; Score 310.4; DB 50; Length 572;
Best Local Similarity 96.1%; Pred. No. 8.4e-86;
Matches 322; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
Oy 1 cccctgattgctcaacaagatgagcattttctcttccatcagtttttactgtg 60
Db 456 cccctgattgctcaacaagatgagcattttctcttccatcagtttttactgtg 397
Oy 61 ttttatcagctaaagtccagagagattgacgtgtgattgagcaaacggagccct 120
Db 396 ttttatcagctaaagtccagagagattgacgtgtgattgagcaaacggagccct 337
Oy 121 gcatgcaagacagagagccacatgccctcacagatgctgtgtgcagagttccagagat 180
Db 336 gcatgcaagacagagagccacatgccctcacagatgctgtgtgcagagttccagagat 277
Oy 181 accttgaccttcccacacagctgcccctgacgtgacctgtgacattaaattcagaa 240
Db 276 accttgaccttcccacacagctgcccctgacgtgacctgtgacattaaattcagaa 217
Oy 241 accttctctcccaaggaattttttctctacacagctgacctgtgttttccagat 300
Db 216 accttctctcccaaggaattttttctctacacagctgacctgtgttttccagat 157

RESULT 6
US-60-181-428-187/c
: Sequence 187, Application US/60181428
: GENERAL INFORMATION:
: APPLICANT: Bonazzi, Vivien
: TITLE OF INVENTION: ISOLATED HUMAN DRUG-METABOLIZING
: TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN
: FILE REFERENCE: DRUG-METABOLIZING PROTEINS, AND USES THEREOF
: CURRENT APPLICATION NUMBER: US/60/181-428
: NUMBER OF SEQ ID NOS: 02-09
: SOFTWARE: FASTSEQ For Windows Version 4.0
: SEQ ID NO 187
: LENGTH: 572
: TYPE: DNA
: ORGANISM: HUMAN
```

```
Matches 323; Conservative 0; Mismatches 0; Indels 0; Gaps 2;
Oy 1 cccctgattgctcaacaagatgagcattttctcttccatcagtttttactgtg 60
Db 1 cccctgattgctcaacaagatgagcattttctcttccatcagtttttactgtg 60
Oy 61 ttttatcagctaaagtccagagagattgacgtgtgattgagcaaacggagccct 120
Db 61 ttttatcagctaaagtccagagagattgacgtgtgattgagcaaacggagccct 120
Oy 121 gcatgcaagacagagagccacatgccctcacagatgctgtgtgcagagttccagagat 180
Db 121 gcatgcaagacagagagccacatgccctcacagatgctgtgtgcagagttccagagat 180
Oy 181 accttgaccttcccacacagctgcccctgacgtgacctgtgacattaaattcagaa 240
Db 181 accttgaccttcccacacagctgcccctgacgtgacctgtgacattaaattcagaa 240
Oy 241 accttctctcccaaggaattttttctctacacagctgacctgtgttttccagat 300
Db 241 accttctctcccaaggaattttttctctacacagctgacctgtgttttccagat 300
Oy 301 ccccaattcatagatcattttt 323
Db 301 ccccaattcatagatcattttt 323

RESULT 4
US-60-313-371-759
: Sequence 187, Application US/6013371
: GENERAL INFORMATION:
: APPLICANT: Ring, Ruijun Z.
: APPLICANT: Malsen, Gareth
: APPLICANT: Tomley, David
: APPLICANT: Morris, MacDonald
: TITLE OF INVENTION: Single Nucleotide Polymorphisms Associated With ADME Genes
: FILE REFERENCE: GR-0013-5 P
: CURRENT APPLICATION NUMBER: US/60/313-371
: NUMBER OF SEQ ID NOS: 08-16
: SOFTWARE: PERL Program
: SEQ ID NO 759
: LENGTH: 323
: TYPE: DNA
: ORGANISM: Homo sapiens
: NAME/REF: misc.feature
: OTHER INFORMATION: GB:HM2C9X05
US-60-313-371-759

Query Match 100.0%; Score 323; DB 64; Length 323;
Best Local Similarity 100.0%; Pred. No. 7.4e-90;
Matches 323; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1 cccctgattgctcaacaagatgagcattttctcttccatcagtttttactgtg 60
Db 1 cccctgattgctcaacaagatgagcattttctcttccatcagtttttactgtg 60
Oy 61 ttttatcagctaaagtccagagagattgacgtgtgattgagcaaacggagccct 120
Db 61 ttttatcagctaaagtccagagagattgacgtgtgattgagcaaacggagccct 120
Oy 121 gcatgcaagacagagagccacatgccctcacagatgctgtgtgcagagttccagagat 180
Db 121 gcatgcaagacagagagccacatgccctcacagatgctgtgtgcagagttccagagat 180
Oy 181 accttgaccttcccacacagctgcccctgacgtgacctgtgacattaaattcagaa 240
Db 181 accttgaccttcccacacagctgcccctgacgtgacctgtgacattaaattcagaa 240
Oy 241 accttctctcccaaggaattttttctctacacagctgacctgtgttttccagat 300
Db 241 accttctctcccaaggaattttttctctacacagctgacctgtgttttccagat 300

RESULT 5
US-60-170-373-1355/c
: Sequence 1355, Application US/60170373
: GENERAL INFORMATION:
: APPLICANT: Bonazzi, Vivien
: TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
: TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS AND USES
: FILE REFERENCE: THERDOP
: CURRENT APPLICATION NUMBER: US/60/170-373
: NUMBER OF SEQ ID NOS: 198-12-13
: SOFTWARE: FASTSEQ For Windows Version 4.0
: SEQ ID NO 1355
: LENGTH: 572
: TYPE: DNA
: ORGANISM: Human
US-60-170-373-1355

Query Match 96.1%; Score 310.4; DB 50; Length 572;
Best Local Similarity 96.1%; Pred. No. 8.4e-86;
Matches 322; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
Oy 1 cccctgattgctcaacaagatgagcattttctcttccatcagtttttactgtg 60
Db 456 cccctgattgctcaacaagatgagcattttctcttccatcagtttttactgtg 397
Oy 61 ttttatcagctaaagtccagagagattgacgtgtgattgagcaaacggagccct 120
Db 396 ttttatcagctaaagtccagagagattgacgtgtgattgagcaaacggagccct 337
Oy 121 gcatgcaagacagagagccacatgccctcacagatgctgtgtgcagagttccagagat 180
Db 336 gcatgcaagacagagagccacatgccctcacagatgctgtgtgcagagttccagagat 277
Oy 181 accttgaccttcccacacagctgcccctgacgtgacctgtgacattaaattcagaa 240
Db 276 accttgaccttcccacacagctgcccctgacgtgacctgtgacattaaattcagaa 217
Oy 241 accttctctcccaaggaattttttctctacacagctgacctgtgttttccagat 300
Db 216 accttctctcccaaggaattttttctctacacagctgacctgtgttttccagat 157

RESULT 6
US-60-181-428-187/c
: Sequence 187, Application US/60181428
: GENERAL INFORMATION:
: APPLICANT: Bonazzi, Vivien
: TITLE OF INVENTION: ISOLATED HUMAN DRUG-METABOLIZING
: TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN
: FILE REFERENCE: DRUG-METABOLIZING PROTEINS, AND USES THEREOF
: CURRENT APPLICATION NUMBER: US/60/181-428
: NUMBER OF SEQ ID NOS: 02-09
: SOFTWARE: FASTSEQ For Windows Version 4.0
: SEQ ID NO 187
: LENGTH: 572
: TYPE: DNA
: ORGANISM: HUMAN
```

US-60-181-428-187

Query Match 96.1% Score 310.4; DB 51; Length 572;
Best Local Similarity 99.4%; Pred. No. 8.4e-86;
Matches 322; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

Qy 1 cccctgaattgtacacaaatgtgcattttttctcttccatcagtttttacttg 60
Db 456 CCCCTGAATGCTACACAAATGTGCCATTTTCTCTTTCCTCATGTTTACTTG 397
|||||

Qy 61 tcttatcagctaaatgcaggagattgaacgtgtgattggcagaaccggagccct 120
Db 396 TCTTATCAGCTAAAGTCCAGAGAGATTGAGCTGTGATTGGCAGAACCGAGCCCT 337
|||||

Qy 121 gcatgcaagcaggagccatgcctacacagatgctgtgtgcagaggtccagat 180
Db 336 GCATGCAAGCAGGAGCCATGCCCTACACAGATGCTGTGTGCACGAGTCCAGAT 277
|||||

Qy 181 accttgaccttctcccaaccagctgccctccatgcagtgacctgtgacataaattcagaa 240
Db 276 ACATTTGACCTTCTCCCAACGAGCTGCCCATGCAGTGCAGCTGTGACATTAAATTCAGAA 217
|||||

Qy 241 acctatcctatcccaagtaattgttttctctacacagatccatgctttccagat 300
Db 216 ACTATCTCTATCCCAAGTAAAGTTTGTCTCTACATGCCAAGCTCACTGTTTCGAAGT 157
|||||

Qy 301 -cccaaatcagatcatctttt 323
Db 156 CCCCAATTCATAGTATCATTTTT 133
|||||

RESULT 7
US-60-181-428-188/c
: Sequence 188, Application US/60181428
: GENERAL INFORMATION:
: APPLICANT: Bonazzi, Vivien
: TITLE OF INVENTION: ISOLATED HUMAN DRUG-METABOLIZING
: TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN
: TITLE OF INVENTION: DRUG-METABOLIZING PROTEINS, AND USES THEREOF
: FILE REFERENCE: CL000227
: CURRENT FILING DATE: US/60/181.428
: CURRENT FILING DATE: 2000-02-09
: NUMBER OF SEQ ID NOS: 40
: SOFTWARE: FASTSEQ for Windows Version 4.0
: SEQ ID NO 188
: LENGTH: 572
: TYPE: DNA
: ORGANISM: HUMAN
US-60-181-428-188

Query Match 96.1% Score 310.4; DB 51; Length 572;
Best Local Similarity 99.4%; Pred. No. 8.4e-86;
Matches 322; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

Qy 1 cccctgaattgtacacaaatgtgcattttttctcttccatcagtttttacttg 60
Db 456 CCCCTGAATGCTACACAAATGTGCCATTTTCTCTTTCCTCATGTTTACTTG 397
|||||

Qy 61 tcttatcagctaaatgcaggagattgaacgtgtgattggcagaaccggagccct 120
Db 396 TCTTATCAGCTAAAGTCCAGAGAGATTGAGCTGTGATTGGCAGAACCGAGCCCT 337
|||||

Qy 121 gcatgcaagcaggagccatgcctacacagatgctgtgtgcagaggtccagat 180
Db 336 GCATGCAAGCAGGAGCCATGCCCTACACAGATGCTGTGTGCACGAGTCCAGAT 277
|||||

Qy 181 accttgaccttctcccaaccagctgccctccatgcagtgacctgtgacataaattcagaa 240
Db 276 ACATTTGACCTTCTCCCAACGAGCTGCCCATGCAGTGCAGCTGTGACATTAAATTCAGAA 217
|||||

Qy 241 acctatcctatcccaagtaattgttttctctacacagatccatgctttccagat 300
Db 216 ACTATCTCTATCCCAAGTAAAGTTTGTCTCTACATGCCAAGCTCACTGTTTCGAAGT 157
|||||

US-60-248-498-247/c

Query Match 96.1% Score 310.4; DB 57; Length 51955;
Best Local Similarity 99.4%; Pred. No. 7.5e-85;
Matches 322; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

Qy 1 cccctgaattgtacacaaatgtgcattttttctcttccatcagtttttacttg 60
Db 8041 CCCCTGAATGCTACACAAATGTGCCATTTTCTCTTTCCTCATGTTTACTTG 7982
|||||

Qy 61 tcttatcagctaaatgcaggagattgaacgtgtgattggcagaaccggagccct 120
Db 7981 TCTTATCAGCTAAAGTCCAGAGAGATTGAGCTGTGATTGGCAGAACCGAGCCCT 7922
|||||

Qy 121 gcatgcaagcaggagccatgcctacacagatgctgtgtgcagaggtccagat 180
Db 7921 GCATGCAAGCAGGAGCCATGCCCTACACAGATGCTGTGTGCACGAGTCCAGAT 7862
|||||

Qy 181 accttgaccttctcccaaccagctgccctccatgcagtgacctgtgacataaattcagaa 240
Db 7861 ACATTTGACCTTCTCCCAACGAGCTGCCCATGCAGTGCAGCTGTGACATTAAATTCAGAA 7802
|||||

Qy 241 acctatcctatcccaagtaattgttttctctacacagatccatgctttccagat 300
Db 7801 ACTATCTCTATCCCAAGTAAAGTTTGTCTCTCATGCTCACTGCAACTCCATGTTTCGAAGT 7742
|||||

Qy 301 -cccaaatcagatcatctttt 323
Db 7741 CCCCAATTCATAGTATCATTTTT 7718
|||||

RESULT 9
US-60-248-498-247/c
: Sequence 9, Application US/60245227
: GENERAL INFORMATION:
: APPLICANT: Beasley, Ellen
: TITLE OF INVENTION: ISOLATED HUMAN PHASE I DRUG-METABOLIZING
: TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN
: TITLE OF INVENTION: DRUG-METABOLIZING PHASE I PROTEINS, AND USES THEREOF
: FILE REFERENCE: CL000910
: CURRENT FILING DATE: 2000-11-15
: NUMBER OF SEQ ID NOS: 254
: SOFTWARE: FASTSEQ for Windows Version 4.0
: SEQ ID NO 24
: LENGTH: 51955
: TYPE: DNA
: ORGANISM: HUMAN
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (1)...(51955)
: OTHER INFORMATION: n = A,T,C or G
US-60-248-498-24

Query Match 96.1% Score 310.4; DB 57; Length 51955;
Best Local Similarity 99.4%; Pred. No. 7.5e-85;
Matches 322; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

Qy 1 cccctgaattgtacacaaatgtgcattttttctcttccatcagtttttacttg 60
Db 8041 CCCCTGAATGCTACACAAATGTGCCATTTTCTCTTTCCTCATGTTTACTTG 7982
|||||

Qy 61 tcttatcagctaaatgcaggagattgaacgtgtgattggcagaaccggagccct 120
Db 7981 TCTTATCAGCTAAAGTCCAGAGAGATTGAGCTGTGATTGGCAGAACCGAGCCCT 7922
|||||

Qy 121 gcatgcaagcaggagccatgcctacacagatgctgtgtgcagaggtccagat 180
Db 7921 GCATGCAAGCAGGAGCCATGCCCTACACAGATGCTGTGTGCACGAGTCCAGAT 7862
|||||

Qy 181 accttgaccttctcccaaccagctgccctccatgcagtgacctgtgacataaattcagaa 240
Db 7861 ACATTTGACCTTCTCCCAACGAGCTGCCCATGCAGTGCAGCTGTGACATTAAATTCAGAA 7802
|||||

Qy 241 acctatcctatcccaagtaattgttttctctacacagatccatgctttccagat 300
Db 7801 ACTATCTCTATCCCAAGTAAAGTTTGTCTCTCATGCTCACTGCAACTCCATGTTTCGAAGT 7742
|||||

Qy 301 -cccaaatcagatcatctttt 323
Db 7741 CCCCAATTCATAGTATCATTTTT 7718
|||||

RESULT 9
US-60-245-227-9/c
: Sequence 9, Application US/60245227
: GENERAL INFORMATION:
: APPLICANT: Beasley, Ellen
: TITLE OF INVENTION: ISOLATED HUMAN PHASE I DRUG-METABOLIZING
: TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN
: TITLE OF INVENTION: DRUG-METABOLIZING PHASE I PROTEINS, AND USES THEREOF
: FILE REFERENCE: CL000910
: CURRENT FILING DATE: 2000-11-03
: NUMBER OF SEQ ID NOS: 129

```

: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 9
: LENGTH: 225532
: TYPE: DNA
: ORGANISM: HUMAN
: NAME/KEY: misc_feature
: LOCATION: (1)---(225532)
: OTHER INFORMATION: n = A,T,C or G
US-60-245-227-9

Query Match          96.1%  Score 310.4; DB 57; Length 225532;
Best Local Similarity 96.1%; Pred. No. 9.3e-82;
Matches 322; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

OY 1 cccctgaattgctacacaaatgagcatttttccctttccatcagttttacttg 60
DB 54498 cccctgaattgctacacaaatgagcatttttccctttccatcagttttacttg 54439

OY 61 tcttatcagctaaagtcacagagagatgaactgtgattgagcagaacccgagccct 120
DB 54438 tcttatcagctaaagtcacagagagatgaactgtgattgagcagaacccgagccct 54379

OY 121 gcacgcagagacagacacacatg-ccttacacagatgctgtgacgaggtccagaga 179
DB 54378 gcacgcagagacagacacacatg-ccttacacagatgctgtgacgaggtccagaga 54319

OY 181 accttgaccttctccaccagcctgccatgcagtgacctgtgacattaaattcaga 240
DB 54318 accttgaccttctccaccagcctgccatgcagtgacctgtgacattaaattcaga 54259

OY 241 actatctctccacaggaagtgattgtttctctctacactgcaactcattttcgaag 300
DB 54258 actatctctccacaggaagtgattgtttctctctacactgcaactcattttcgaag 54199

OY 301 tcccaaatccatagatcatttt 323
DB 54198 tcccaaatccatagatcatttt 54175

RESULT 10
US-60-213-795-56
: Sequence 56, Application US/60213795
: GENERAL INFORMATION:
: APPLICANT: Beasley, Ellen
: TITLE OF INVENTION: ISOLATED HUMAN PHASE I DRUG-METABOLIZING
: TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN
: TITLE OF INVENTION: DRUG-METABOLIZING PHASE I PROTEINS, AND USES THEREOF
: FILE REFERENCE: CL000705
: CURRENT APPLICATION NUMBER: US/60/213,795
: CURRENT FILING DATE: 2000-06-23
: NUMBER OF SEQ ID NOS: 267
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 56
: LENGTH: 10097
: TYPE: DNA
: ORGANISM: HUMAN
: NAME/KEY: misc_feature
: LOCATION: (1)---(10097)
: OTHER INFORMATION: n = A,T,C or G
US-60-213-795-56

Query Match          92.7%  Score 299.4; DB 54; Length 10097;
Best Local Similarity 99.1%; Pred. No. 9.3e-82;
Matches 322; Conservative 0; Mismatches 1; Indels 2; Gaps 2;

OY 1 cccctgaattgctacacaaatgagcatttttccctttccatcagttttacttg 60
DB 5009 cccctgaattgctacacaaatgagcatttttccctttccatcagttttacttg 5068

OY 61 tcttatcagctaaagtcacagagagatgaactgtgattgagcagaacccgagccct 120
DB 5059 tcttatcagctaaagtcacagagagatgaactgtgattgagcagaacccgagccct 5128

OY 121 gcacgcagagacagacacacatg-ccttacacagatgctgtgacgaggtccagaga 179
DB 5129 gcacgcagagacagacacacatg-ccttacacagatgctgtgacgaggtccagaga 5188

OY 180 tacttgaccttctccaccagcctgccatgcagtgacctgtgacattaaattcaga 239
DB 5189 tacttgaccttctccaccagcctgccatgcagtgacctgtgacattaaattcaga 5248

OY 240 actatctctccacaggaagtgattgtttctctctacactgcaactcattttcgaag 299
DB 5249 actatctctccacaggaagtgattgtttctctctacactgcaactcattttcgaag 5308

OY 300 tcccaaatccatagatcatttt 323
DB 5309 tcccaaatccatagatcatttt 5333

US-60-213-795-56
: Sequence 56, Application US/60213795
: GENERAL INFORMATION:
: APPLICANT: Beasley, Ellen
: TITLE OF INVENTION: ISOLATED HUMAN PHASE I DRUG-METABOLIZING
: TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN
: TITLE OF INVENTION: DRUG-METABOLIZING PHASE I PROTEINS, AND USES THEREOF
: FILE REFERENCE: CL000705
: CURRENT APPLICATION NUMBER: US/60/213,795
: CURRENT FILING DATE: 2000-06-23
: NUMBER OF SEQ ID NOS: 267
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 56
: LENGTH: 10097
: TYPE: DNA
: ORGANISM: HUMAN
: NAME/KEY: misc_feature
: LOCATION: (1)---(10097)
: OTHER INFORMATION: n = A,T,C or G
US-60-213-795-56

Query Match          92.7%  Score 299.4; DB 54; Length 10097;
Best Local Similarity 99.1%; Pred. No. 9.3e-82;
Matches 322; Conservative 0; Mismatches 1; Indels 2; Gaps 2;

OY 1 cccctgaattgctacacaaatgagcatttttccctttccatcagttttacttg 60
DB 5009 cccctgaattgctacacaaatgagcatttttccctttccatcagttttacttg 5068
```

[illegible]

```

      298 agt-cccacattcatgatcatttt 323
          ||| ||||| ||||| ||||| |||||
DB    2326 agtccccaattcatgatcattt 2352

RESULT 14
US     207-207-211-21
SEQUENCE 21, Application US/60207211
GENERAL INFORMATION:
APPLICANT: Beasley, Ellen
TITLE OF INVENTION: ISOLATED HUMAN PHASE I DRUG-METABOLIZING
PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN
TITLE OF INVENTION: ISOLATED HUMAN PHASE I DRUG-METABOLIZING
PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN
PUBLICATION NUMBER: US/60-207-211
CURRENT APPLICATION NUMBER: US/60/207-211
CURRENT FILING DATE: 2000-05-26
NUMBER OF SEQ ID NOS: 74
SOFTWARE: FaalSeq for Windows Version 4.0
SEQ ID NO 21
LENGTH: 3557
ORGANISM: HUMAN
US-60-207-211-21

```

Query Match	88.0%	Score	284.2	DB	53:	Length	3557:
Best Local Similarity	97.9%	Pred. No.	3.1e-77:				
Matches	320:	Conservative	0:	Mismatches	3:	Indels	4:
Gaps	3:						
Oy	1	ccccgaattgctacacaaatgacctttttctcttttccatcagatttttacttgta	60				
Db	2026	ccccgaattgctacacaaatgacctttttctcttttccatcagatttttacttgta	60				
Oy	61	ttcttatcagctaaagctccgaggaagattgaacgttgatggcgaaaccc-ggaagccc	118				
Db	2086	ttcttatcagctaaagctccgaggaagattgaacgttgatggcgaaacccgagagccc	118				
Oy	119	ctgcatgcacagcaggagccatg-ccctacacagatgctgttggtgacaggttcagga	177				
Db	2146	ctgcatgcacagcaggagccatg-ccctacacagatgctgttggtgacaggttcagga	177				
Oy	178	atacattctgaccttctccacacagctgcccctcagatgacctctgacattcaaatcca	237				
Db	2206	atacattctgaccttctccacacagctgcccctcagatgacctctgacattcaaatcca	237				
Oy	238	gatacatatctatcccaaggaagattgtttctctcagctgaacatcatattttcaga	295				
Db	2266	gatacatatctatcccaaggaagattgtttctctcagctgaacatcatattttcaga	295				
Oy	2266	gatacatatctatcccaaggaagattgtttctctcagctgaacatcatattttcaga	2325				

Search completed: April 19, 2002, 11:07:02
Job time: 13522 sec

Result	No.	Score	Query	Match	Length	DB	ID	Description
	1	1739.6	99.6	1746	1	US-08-201-118-2		Sequence 2, Appli
	2	1739.6	99.6	1746	2	US-08-238-8218-2		Sequence 2, Appli
	3	1739.6	99.6	1746	5	PCT-US95-05744-2		Sequence 2, Appli
	4	1534.6	87.9	1852	1	US-08-201-118-10		Sequence 10, Appl
	5	1534.6	87.9	1852	2	US-08-238-8218-10		Sequence 10, Appl
	6	1534.6	87.9	1852	5	PCT-US95-05744-10		Sequence 10, Appl
	7	1529.8	87.6	1854	1	US-08-201-118-4		Sequence 4, Appli
	8	1529.8	87.6	1854	2	US-08-238-8218-4		Sequence 4, Appli
	9	1529.8	87.9	1852	5	PCT-US95-05744-4		Sequence 4, Appli
	10	1432.8	82.3	1892	1	US-08-201-118-14		Sequence 14, Appl
	11	1436.4	82.3	1892	2	US-08-238-8218-14		Sequence 14, Appl
	12	1436.4	82.3	1892	5	PCT-US95-05744-14		Sequence 14, Appl
	13	1373.6	78.7	1591	2	US-08-134-981E-3		Sequence 3, Appli
	14	1276.6	73.1	1419	2	US-08-194-981E-4		Sequence 4, Appli
	15	1264.6	72.4	2009	1	US-08-201-118-6		Sequence 6, Appli
	16	1264.6	72.4	2009	2	US-08-238-8218-6		Sequence 6, Appli
	17	1264.6	72.4	2009	5	PCT-US95-05744-6		Sequence 6, Appli
	18	1264.6	72.4	2258	1	US-08-201-118-12		Sequence 12, Appl
	19	1264.6	72.4	2258	2	US-08-238-8218-12		Sequence 12, Appl
	20	1264.6	72.4	2258	5	PCT-US95-05744-12		Sequence 12, Appl
	21	1170.0	67.0	1829	1	US-08-201-118-8		Sequence 8, Appli
	22	1170.0	67.0	1829	2	US-08-238-8218-8		Sequence 8, Appli
	23	671.2	35.9	1863	1	US-08-194-981E-1		Sequence 1, Appli
	24	671.2	35.9	1863	2	US-08-238-8218-1		Sequence 1, Appli
	25	611.4	35.0	1419	2	US-08-134-981E-2		Sequence 2, Appli
	26	468.6	26.8	1737	2	US-08-750-703-8		Sequence 8, Appli
	27	465.6	26.7	1740	2	US-08-750-703-1		Sequence 1, Appli

SUMMARIES

[illegible]


```

1141 ACCTCATCCCAAGGACACACATATATCTCCCTCCTCCTCTGCTGCTACATGACACA 1200
1201 agaatttcccaacagagagtggtgacccctgctcacttcttgatgaagtggaatt 1260
1201 AGAATTTCACACCCAGAGATGTTGACCTGCTGCTCTCTGATGAGGTGGAAT 1260
1261 ttgaagaagaactactctgactcttctcagcagaagaacagatttgctggagaag 1320
1261 TTGAAGAAGTACTACTCTGCTCTCTCAGCAGAAACGATTTGCTGGAGAG 1320
1321 gctggcgcgagtgagtggttttctcctgactcttcttctacagaactttaacctga 1380
1321 GCTGGCGCGATGGAGCTGTTTATCTCTGACCTCATTTTACAGAAGTTTAACTCTGA 1380
1381 aatctctgattgacccaagacactctgtgtcgaatgattgattctgtg 1440
1381 AATCTCTGATTGACCCAAGACCTGTGACACACTCTGTGTGATGATTTGCTCTG 1440
1441 tccgcgcctctcagctgactctcactctcagcagaacacagatgctgactga 1500
1441 TCCGCCTCTCTATCAGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1500
1501 ctctgctgctgctcagcagctctctctcctgctgctgctgctgctgctgctgct 1560
1501 CTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1560
1561 tctctgacgcgtctctcactctctctctctctctctctctctctctctctctct 1620
1561 TCTCTGACGCTCTCTCAGATTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1620
1621 catcaaaagctctcagcagaataatctctctctctctctctctctctctctctct 1680
1621 CATCAAAAGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1680
1681 attgagtgacataatgctgactctctctctctctctctctctctctctctctct 1740
1681 ATTAGTGACATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1740
1741 aataga 1746
1741 AATAGA 1746

RESULT 4
US-08-201-118-10
US-08-201-118-10 Application US/08201118
Patent No. 5786191
GENERAL INFORMATION:
APPLICANT: GOLDSTEIN, Joyce A.
APPLICANT: ROMKES-SPARKS, Marjorie
TITLE OF INVENTION: CLONING AND EXPRESSION OF COMPLEMENTARY
TITLE OF INVENTION: DNAs FOR MULTIPLE MEMBERS OF THE HUMAN CYTOCHROME P450 2C
TITLE OF INVENTION: SUBFAMILY
NUMBER OF SEQUENCES: 44
CORRESPONDENCE ADDRESS:
APPLICANT: Townsend and Townsend Kourile and Crew
STREET: 379 Lyndon Avenue
CITY: Palo Alto
STATE: California
COUNTRY: US
ZIP: 94301
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE RELEASE: Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/201.118
FILING DATE: 22-FEB-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/864,962
```

```

FILING DATE: 09-APR-1992
ATTORNEY/AGENT INFORMATION:
NAME: 'Liebeschuetz, Joe
REGISTRATION NUMBER: 37,505
REFERENCE/DOCKET NUMBER: 15280-192-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 326-7240
FACSIMILE: (415) 326-7240
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 1852 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-201-118-10

Query Watch      87.9%; Score 1534.6; DB 1; Length 1852;
Best Local Similarity 92.9%; Pred. No. 0;
Matches 1637; Conservative 0; Mismatches 109; Indels 17; Gaps 2;

Cy 1 cttcaatgagtcctttgtgctgctgctgctgctgctgctgctgctgctgcttcca 60
Db 6 CTTCATAGATGATCTCTGTGGCTGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 65
Cy 51 tctggagcagagctctggagagaaactcctctggtcccaactcctctccagtgga 120
Db 56 TCTGGAGCAGAGCTCTGGAGAGAAACTCTCTCTGGGCCCTCTCTCCGCTGA 125
Cy 121 ttggaatatctcacaagatagatatgaagatgctcagcaaatctcttaacaaatcttcaa 180
Db 126 TTGGAATATCTTACAGATAGTATTAAAGACATCAGCAATCTTAAACCAATCTCTCAA 185
Cy 181 aaactatgagctctgttccactctgtatttggcctggagcagatggtgctgcatg 240
Db 186 AGACTATAGGCTGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 245
Cy 241 aatgaagtgaggagagagcctgattgctcttggagagagagcttcttggagagagcc 300
Db 245 GATATGACGACGTGAAGAGAGCCCTGATGATCTTGGAGAGAGAGTCTTCTGGAGAGGCA 305
Cy 301 atttcccaactgagcagaagcagaagatttggaatcgcttttcccaactgagaaga 360
Db 306 TTTTCCCTAGCTGAGAGAGCTACAGAGATTTGGATTTGTTTACACATGGAGAGA 365
Cy 351 aatggaagagatcgacattcttccctcactaacgcgcggaatttggagtggaaga 420
Db 356 AATGGAAGAGATCTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 425
Cy 421 ggaacatgagagcgtcttcaagagagagccctcctcctctgctggagagtgagaagaa 480
Db 426 GGAGCATGAGAGCGGTGTTCAAGAGAGAGCGCGCTGCTTGTGGAGAGTTGAGAAA 485
Cy 481 ccaagctcaccctgagatcccaattctcctggtgctgctcctcctcgaatgatct 540
Db 486 CGAGGCTCACTCTGTATCCCACTTTCATCTCTGGCTGTGCTCTGCTGCTGCTGCTG 545
Cy 541 gctcattatttccagaagcgttctgattataaagatcacgaattcttaacttgatgg 600
Db 546 GCTCATATTITTCATAAAGCTTTTGATTATAAAGATCAGCAATTTCTTAACCTAATG 605
Cy 601 aaaaattgaataaacatcagagtgtaagcccccttggatccagatagcaataatt 660
Db 606 AAGATTGATGAACATCAAGATTTGACAGCGCTGATCCAGATCTGCATATATT 665
Cy 661 ttccactatcattgattttcccggaacccataaacaatttacttaaaacctgttt 720
Db 666 TTCTCTATCATGATGATCTTCCCGGGAAGTCAACAAATTTACTTAAAGAGTGTCT 725
Cy 721 ttatgaaagtgatatttggagaaataaagaacacacacacacacacacacacac 780
Db 726 TTATGAAGAGTTATTTTGGAAAAGTAAAGAACACCAAGATCATATGACATGACAA 785
```


Oy 241 gatatgaagtgtaaggaagccctgattgacttggagaggtttttctggaagagcc 300
Db 246 CATATGAGGACGACGAGGAGCCCTGATTGATCTTGGAGAGGATTTTCTGAGAGGCA 305
Oy 301 atttcccaactggctgaagagagcttaacagagatttggaaactgtttccagcaatgggaaga 360
Db 306 TTTTCCCACTGGCTGAAGAGGCTTACAGAGGATTTGGAAATGTTTTCAGCAATGGAAAGA 365
Oy 361 gataagagagatcggagtttctctccatcagagcttcggaaattttggagaggaaga 420
Db 366 ATGAGAGGAGATCGGAGTTTCTCCCTGATGAGCCTGGGAATTTTGGATGGGGAAGA 425
Oy 421 ggaagatggagagctgttcaagagagagcccgctgcttggagagagtggaagaaa 480
Db 426 GGAGCATTTGAGGAGCCGCTGTCAAGAGAGAGCCGCTGCTTGTGGAGGAGTTTGAGAAAAA 485
Oy 481 ccaaggttcaacccctgataccttcaacatttcaacttgggtgtgctccctgcaatgtatct 540
Db 486 CCAAGGCTCCACCTGTGATCCCACTTTCATCTCGGGGTGTGCTCCCTGCAATGTATCT 545
Oy 541 gctccatttttccagagcttgattataaagatacgaatttttcaactgaatg 600
Db 546 GCTCCATTTTTCATTAAGCTTTTGATTATAAAGATCAGCAATTTCTTAACTTAATGG 605
Oy 601 aaaaattgaatgaatacatcgagattgtaagcaaccccttgatccagatagcaataatt 660
Db 606 AAAAAATTGAATGAATCAAGATTTTGAAGAGAGCCCTGCTGATCCAGATCTGCATAAT 665
Oy 661 tcccaactatctgatttccagagagagcccttaacaaacttaacttaaaacttactt 720
Db 666 TTTCTCTATCATGATTACTTCCCGGGAACTCACACAAATTAAGTAAAAAGCTTGCT 725
Oy 721 ttatgaaagtatatttggagaagttaaagaacaccccaagatcgatggscatcaaca 780
Db 726 TTAAGAAAGTTATATTTTGAAGAAAGTAAAGAAACCAACCAAGAAATCAATGGACATGA 785
Oy 781 cccctgggttattatgttctctgtcaaatggagagagaaagaaacaaac 840
Db 786 ACCTCAGGACTTTATGATTGCTCTGATCAAAATGGAGAGGAAAGCAACCAAC 845
Oy 841 agtctgaattcaactattgaaacttggttaactcactgcagctgacttctgagctggga 900
Db 846 CATTCTGAATTTACTTGAAGCTTGGAAACACTGCACTTGACTTGTGAGCTGGGA 905
Oy 901 cagagcaacaaagcaacccctgagatagtctctctctcctgcagagcccaagg 960
Db 906 CAGAGCAACAGCAACACCCCTGAGATATCTCTCTCTCTGAGACCAACCAAGG 965
Oy 961 tcaagctaaagtcaagagagattcgactgctcattgagaaacccgctctga 1020
Db 966 TCACAGCTTAAAGTCCAGGAAGATTGAAGCTGTGATTGGCAGAAACCGGAGCCCTGCA 1025
Oy 1021 tgaacagaggggacacatgcctctacagatgctgtgtgtgcagagtcagagataca 1080
Db 1026 TGCAGAGAGGAGCCATGCT 1085
Oy 1081 tgaactctcccaagcctgcccctgagctgagctgagctgagcttaaatcaaaact 1140
Db 1086 TTGAAGCTTCTCCCAACACCGCTGCCCATGCACTGCTGTGACATTAAATTCAGAACT 1145
Oy 1141 acctcttcccaaggcacaacacattataactctctctctctctctctctctctctct 1200
Db 1146 ATCTCATTTCCCAAGGCGCAACCATTAATTTCCCTGACTTCTGTGTACATGACACA 1205
Oy 1201 agaatctcccaacccagagatgtttgacctgtcactcttctgagatgaagtgaatt 1260
Db 1206 AAGATTTTCCCAACCCAGAGATGTTTGAACCTCATCACTTCTGAGTGAAGTGGCAAT 1265
Oy 1261 taaagaaaglaactacttctctctctctctcagcagagaaacagatttggggagag 1320
Db 1266 TTAAGAAAGTAAATTAATCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1325

RESULT 6
PCT-US95-05744-10
: Sequence 10, Application PC/TUS9505744
: GENERAL INFORMATION: JIN, Joyce A.
: APPLICANT: ROMES-SPARKS, Marjorie
: APPLICANT: ROMES-SPARKS, Marjorie
: APPLICANT: DE MORAIS, Sonia M.F.
: TITLE OF INVENTION: CLONING, EXPRESSION AND DIAGNOSIS OF HUMAN
: TITLE OF INVENTION: CYTOCHROME P450 2C19: THE PRINCIPAL DETERMINANT
: NUMBER OF SEQUENCES: 61
: CORRESPONDENCE ADDRESS: 61
: ADDRESSEE: Townsend and Townsend Kourile and Crew
: STREET: 175 Lytton Avenue
: CITY: Palo Alto
: STATE: California
: COUNTRY: US
: ZIP: 94301
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent, Release #1.0, Version #1.25
: CUSTODIAN: JIN, Joyce A.
: APPLICATION NUMBER: PCT/US95/05744
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION NUMBER: US 08/238,821
: FILING DATE: 05-MAY-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/201,118
: FILING DATE: 08-FEB-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/864,962
: FILING DATE: 09-APR-1992
: ATTORNEY/AGENT INFORMATION:
: NAME: Dow, Karen B.
: REGISTRATION NUMBER: 29,684
: REFERENCE/DOCKET NUMBER: 15280-192-1-1

```

:
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 326-2400
: TELEFAX: (415) 326-2422
: INFORMATION FOR SEQ ID NO: 10:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1852 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: PRT: US95-05744-10
:

```

Query Match	87.9%	Score 1534.6;	DB 5;	Length 1852;
Best Local Similarity	92.9%	Pred. No. 0;		
Matches 1637;	Conservative	0;	Mismatches 109;	Indels 17; Gaps 2;

[illegible][illegible]

```

: RESULT 7
: US 08/201-118-4
: : Subject 4, Application US/08201118
: : Patent No. 5786191
: : GENERAL INFORMATION:
: : APPLICANT: GOLDSTEIN, JOYCE A.
: : APPLICANT: ROMKES-STARKS, MATTHEW
: : TITLE OF INVENTION: CLONING AND EXPRESSION OF COMPLEMENTARY

```



```
Oy 541 gctccattttccgaagacgtttcagattataaagatcacgaattcttcaattgatag 600
Db 548 gctccattttttccgaagacgtttcagattataaagatcacgaattcttcaattgatag 607
Oy 601 aaaaatgaagaaacacagatgtgaagccccccttgatccagatgacgaataatt 950
Db 608 aaaaatgaagaaacacagatgtgaagccccccttgatccagatgacgaataatt 957
Oy 661 tccccatcatgatatttatttccgggaacccatacaaatctactaaacatttgctt 720
Db 668 tttctctcatcattgatttacttccgggaacccatacaaatctactaaacatttgctt 727
Oy 721 tttatggaagatgattttggaagaaatgaagaaacacgaatcgatgacatcaaa 780
Db 728 tttatggaagatgatttttggaaaagtgaagaaacacgaatcgatgacatcaaa 787
Oy 781 accctcggaattatttattgcttcctgataaaatggagaagaaagcaaaccaac 840
Db 788 accctcggaattatttattgcttcctgataaaatggagaagaaagcaaaccaac 847
Oy 841 agctcatcctcatttataaactcgaatcgaatcgaatcgaatcgaatcgaatcga 900
Db 848 catctgaatttacttattgaagcttggaaaacacacgaatcgaatcgaatcga 907
Oy 901 cagagacaacagacacccctgagatgcttctctctctctctctctctctctct 950
Db 908 cagagacaacagacacccctgagatgcttctctctctctctctctctctctct 957
Oy 961 tccagctaaagtcagagagatgagagatgagagatgagagatgagagatgagag 1020
Db 968 tccagctaaagtcagagagatgagagatgagagatgagagatgagagatgagag 1027
Oy 1021 tgcagcagagagagagagagagagagagagagagagagagagagagagagag 1080
Db 1028 tgcagcagagagagagagagagagagagagagagagagagagagagagagag 1087
Oy 1081 tgcagcagagagagagagagagagagagagagagagagagagagagagagag 1140
Db 1088 tgcagcagagagagagagagagagagagagagagagagagagagagagagag 1147
Oy 1141 acctcatttcccaagggcacaacacacacacacacacacacacacacacac 1200
Db 1148 acctcatttcccaagggcacaacacacacacacacacacacacacacacac 1207
Oy 1201 aagaatttcccaagggcacaacacacacacacacacacacacacacacacac 1250
Db 1208 aagaatttcccaagggcacaacacacacacacacacacacacacacacacac 1257
Oy 1261 ttaagaagaatgaatcctcctcctcctcctcctcctcctcctcctcctcct 1320
Db 1268 ttaagaagaatgaatcctcctcctcctcctcctcctcctcctcctcctcct 1327
Oy 1321 gctcggccctgagagagagagagagagagagagagagagagagagagagag 1380
Db 1328 gctcggccctgagagagagagagagagagagagagagagagagagagagag 1387
Oy 1381 aatctgatgagagagagagagagagagagagagagagagagagagagagag 1440
Db 1388 aatctgatgagagagagagagagagagagagagagagagagagagagagag 1447
Oy 1441 cccagccctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 1500
Db 1448 cccagccctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 1507
Oy 1501 cctcgtgctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 1553
Db 1508 cctcgtgctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 1567
Oy 1554 tgaatgctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 1613
Db 1568 tgaatgctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 1627
```

Oy 1614 caccctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 1663
Db 1628 caccctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 1687
Oy 1664 taccctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 1723
Db 1688 taccctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 1747
Oy 1724 taccctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 1746
Db 1748 taccctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 1770

RESULT 10
US-08-201-118-14
: Sequence 14, Application US/08201118
: Patent No. 5786191
: GENERAL INFORMATION:
: APPLICANT: GOLDSTEIN, Joyce A.
: INVENTOR: GOLDSTEIN, Joyce A.
: TITLE OF INVENTION: CLONING AND EXPRESSION OF COMPLEMENTARY
: TITLE OF INVENTION: DNAs FOR MULTIPLE MEMBERS OF THE HUMAN CYTOCHROME P450 2C
: TITLE OF INVENTION: SUBFAMILY
: NUMBER OF SEQUENCES: 44
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Townsend and Townsend Kourile and Crew
: STREET: 379 Lytton Avenue
: CITY: Palo Alto
: STATE: California
: COUNTRY: US
: ZIP: 94301
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent in Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/201.118
: FILING DATE: 09-APR-1992
: CLASSIFICATION: 4-435
: PRIORITY INFORMATION:
: APPLICATION NUMBER: US 07/864,962
: FILING DATE: 09-APR-1992
: ATTORNEY/AGENT INFORMATION:
: NAME: Liebeschuetz, Joe
: REGISTRATION NUMBER: 37,505
: REFERENCE/DOCKET NUMBER: 13280-192-1
: TELECOMMUNICATIONS INFORMATION:
: TELEPHONE: (415) 326-2400
: TELEFAX: (415) 326-2422
: INFORMATION FOR SEQ ID NO: 14:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1892 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: RECOMBINATION: YES
US-08-201-118-14

Query Match 82.3% Score 1436.4 DB 1: Length 1892:
Best Local Similarity 86.5% Pred. No. 0:
Matches 1525: Conservative 0: Mismatches 221: Indels 17: Gaps 2:
Oy 1 cttcaatggatccttttgatgctccttgatcctcctcctcctcctcctcctcctc 60
Db 38 cttcaatggatccttttgatgctccttgatcctcctcctcctcctcctcctcctc 97
Oy 61 tctgaagacagacactcctgagagagagagagagagagagagagagagagagagag 120
Db 98 tctgaagacagacactcctgagagagagagagagagagagagagagagagagagag 157

121 tgggaatctacagatagatataagagatcagcaaatcttaaccaatctcga 180
Db 158 TTGAAATATCTACAGATAGATNTAAGGACATCAGCAAAATCTTAACTTCTCA 217
Oy 181 aaatctatgacctctggttcaactctgtatttggcctggagcagatgggtgctcga 240
Db 218 AGTCTATGGCCCTGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 277
Oy 241 gatatgaatgtggaag 300
Db 278 GATATGAAGCAGTGAAGGAGCCCTGATTGATCTGAGAGAGAGAGTCTTCTGGA 337
Oy 301 atttcccaatggtctgaagagagcctgaagagagatggaatcttcaagagagaga 360
Db 338 MTTCCTCCAGCTGGCTGAAGAGNTAACAGAGNTGGAAATCGTTTTCAGCAATGA 397
Oy 361 gatggaagagagattcggcgttctctctctctctctctctctctctctctctct 420
Db 398 GATGAGAGAGAGATCCGGGTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 457
Oy 421 gtcgcatct 480
Db 458 GGAGCATGAGGAGCCGTGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 517
Oy 481 caagagctcaacctgtatctcaactctcaactctctctctctctctctctctct 540
Db 518 CCAAGGCTCAACCTGTATGCTCACTTCACTCTGGGCTGTCTCTCTCTCTCTCT 577
Oy 541 gctcattct 600
Db 578 GCTCCCTATTATTCATAAGCAGTTCATTATTAAGATCAGNAATTCATTACCTG 637
Oy 601 aaaaatgaaatgaatcagagatgtaagagagagagagagagagagagagagag 660
Db 638 AAAAAATTAAATGAATCAGAGATCTCGAGCAGAGAGAGAGAGAGAGAGAGAGAG 697
Oy 661 tcccaactcattgatttattccggagagagagagagagagagagagagagagag 720
Db 698 TCTCTCTATGATGATATTTCTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 757
Oy 721 tctggaagagagatttggagagagagagagagagagagagagagagagagagag 780
Db 758 TATGAAAAGTTATATTGGAGAAAGTAAAGAGACCAAGAAATCANTGACATGA 817
Oy 781 accctggagcttattgattctctctctctctctctctctctctctctctctctct 840
Db 818 AKCTCGGAGCTTTATGATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 877
Oy 841 agctgcatct 900
Db 878 AGTCTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 937
Oy 901 cagagacacagcacaacccctgagatgctctctctctctctctctctctctct 950
Db 938 CAGAGACAG 997
Oy 961 tccagctgaagtcag 1020
Db 998 TCAAGCTAAAGTCAGAGAGATGAGAGTATTTGGAGAGAGAGAGAGAGAGAGAG 1057
Oy 1021 tgcagcagagggccacatgccctacacagatgctggtgagcagaggtccagaata 1080
Db 1058 TCGAG 1117
Oy 1081 tgaacctatccccacagcctgccccatgagtgacctgagctgagcttaattcaga 1140
Db 1118 TTGACTTCTCCACAGCTGCCCCCTGAGCTGAGCTGAGCTGANNITAAATTCAGA 1177
Oy 1141 acctcaatcccaagggcacaacatattactctctctctctctctctctctct 1200
Db 1178 ACTCATMCCAAGGGCACACCATANTAACTGCTGCTGCTGCTGCTGCTGCTG 1237
Oy 1201 aagaatttcccaacccagagagatttgaacctctgctctctctctctctctct 1250

1238 AAGAAATTCOCACCCAGAGATGTTGACCCNGKACTTTCGTATNANNNGTGGCA 1297
Oy 1261 ttaagaaagaaactactctctctctctcagcagaaacagagatttctgtggagag 1320
Db 1298 ITAAGAAAGTACTTCTATGCTTCTCAGCAGAAACGAGATTTGTGGAGANG 1357
Oy 1321 gcttggcgcagatggagctgttttattctctgacctcttctacagaaacttaac 1380
Db 1358 GCTTGGCCCGCATGAGCTGTTTTATTTCTGTGCCNCATTTACAGAGATTTAC 1417
Oy 1381 actctgtatgagcagcagcctctgacaaactctgtgtcagatgagattctctct 1440
Db 1418 AATCTGCTGTGACCCAAAGACCTTGACACCACTCCAGTTGCMAATGGATTG 1477
Oy 1441 tccgcctctctacagctgctgtctctctctgagagagacagatgctgctgctg 1500
Db 1478 TGCNCCTCTTACACCTGTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1537
Oy 1501 ctctctgtct 1553
Db 1538 CTCTCTGCTGTCNCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1597
Oy 1554 tgaagctctctctgacctctctctctctctctctctctctctctctctctct 1613
Db 1598 TTTGTCNNNTTTCATGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1657
Oy 1614 cagct 1663
Db 1658 NANNNNCAATTCAGAGAGTGGNNNTTNNNNNNNNNNNNNNNNNNNNNNNN 1717
Oy 1664 tactctataagattgagctgagcagcagcagcagcagcagcagcagcagc 1723
Db 1718 TNCNT 1777
Oy 1724 ttaacatattattataalaga 1746
Db 1778 TIGSNANNNTTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNT 1800

RESULT 11
US-08-238-821B-14
: Sequence 14, Application US/08238821B
: Patent No. 5912120
: GENERAL INFORMATION:
: APPLICANT: GOLDSTEIN, Joyce A.
: APPLICANT: DE MORAS, Sonia M.F.
: TITLE OF INVENTION: CYTOCHROME P450 2C19, THE PRINCIPAL DETERMINANT OF S-
: TITLE OF INVENTION: METOPROLOL METABOLISM
: NUMBER OF SEQUENCES: 61
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Townsend and Townsend and Crew LLP
: STREET: Two Embarcadero Center, 8th floor
: CITY: San Francisco
: STATE: California
: COUNTRY: US
: ZIP: 94111
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/238-821B
: FILING DATE: 06-MAY-1994
: PRIORITY DATA:
: PRIORITY NUMBER: 435
: APPLICATION NUMBER: US 08/201,118
: FILING DATE: 22-FEB-1994
: CLASSIFICATION: 435
: PRIORITY APPLICATION DATA:


```
568 AAGATTATATTTTGGAAAAAGTAAAGAACACACCAAGCAATCAATGCATGACATGAACACCTC 727
Oy 787 ggaacttatgatctctctctctctctctctctctctctctctctctctctctctctctct 846
Db 728 AGAGCTTATGATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 787
Oy 847 aactctctctctctctctctctctctctctctctctctctctctctctctctctctctctct 906
Db 788 AATTTATTTTATTTTAACTTGAACCTTGAACCTTGAACCTTGAACCTTGAACCTTGAACCT 847
Oy 907 caaagacacacacacacacacacacacacacacacacacacacacacacacacacacacacac 966
Db 848 CGAACGACCAACCTCTGATATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 907
Oy 967 cttaagctcgaagagatgaaactgtgactgtgactgtgactgtgactgtgactgtgactgtg 1026
Db 908 CTAAAGTCCAGAGAGATGACAGTGTGATGACAGTGTGATGACAGTGTGATGACAGTGTGAT 967
Oy 1027 acagagacacacacacacacacacacacacacacacacacacacacacacacacacacac 1086
Db 968 ACAGAGACCAACCTCTGATATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1027
Oy 1087 tcaatccacacacacacacacacacacacacacacacacacacacacacacacacacacac 1146
Db 1028 TTCTCCACCAACCTCTGATATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1087
Oy 1207 tcccaacacacacacacacacacacacacacacacacacacacacacacacacacacacac 1266
Db 1148 TTCCACCAACCTCTGATATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1207
Oy 1267 aaagtaactctctctctctctctctctctctctctctctctctctctctctctctctctct 1326
Db 1208 AAGTAAATATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1267
Oy 1327 ccagatgagagctctctctctctctctctctctctctctctctctctctctctctctct 1386
Db 1268 CGGCATGAGCTGTGTTTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1327
Oy 1387 tgaatgacacacacacacacacacacacacacacacacacacacacacacacacacacac 1446
Db 1328 TGGTTGACCCAAAGACCTTGACACCTTGACACCTTGACACCTTGACACCTTGACACCTTG 1387
Oy 1447 cctctctctctctctctctctctctctctctctctctctctctctctctctctctctct 1478
Db 1388 CTTCTACCACTGTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1419

RESULT 15
US-08-201-118-6
: Sequence 6, Application 05/08201118
: Patent No. 5786191
: GENERAL INFORMATION:
: APPLICANT: ROHMES-SPARES, Marjorie
: TITLE OF INVENTION: CLONING AND EXPRESSION OF COMPLEMENTARY
: TITLE OF INVENTION: DNAs FOR MULTIPLE MEMBERS OF THE HUMAN CYTOCHROME P450 2C
: NUMBER OF SEQUENCES: 44
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Townsend and Townsend Kourile and Crew
: STREET: 379 Lytton Avenue
: CITY: Palo Alto
: STATE: California
: COUNTRY: US
: ZIP: 94301
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
```

```
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/201-118
FILING DATE: 22-FEB-1994
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/884,962
FILING DATE: 09-APR-1992
NAME/AGENT INFORMATION:
NAME: ROHMES-SPARES, Marjorie
REGISTRATION NUMBER: 37,505
REFERENCE/DOCKET NUMBER: 15280-192-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 326-2400
TELEFAX: (415) 326-2422
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 2009 base pairs
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-201-118-6

Query Match 72.4% Score 1264.6 DB 1: Length 2009;
Best Local Similarity 84.4% Pred. No. 0;
Matches 1489; Conservative 0; Mismatches 254; Indels 22; Gaps 5;

Oy 1 cttcaatgagatcttctgtctctgtctctctctctctctctctctctctctctctctctca 60
Db 195 CTTCAATGATCTCCAGCTGTGGCTGTGGCTGTGGCTGTGGCTGTGGCTGTGGCTGTGGCT 254
Oy 61 tctgagacagactctgagagagagaaactctctctctctctctctctctctctctctctctca 120
Db 255 TCTGAGGACAGCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 314
Oy 121 tgaatataatctacagatagatataagagatctcaaatctcaaatctcaaatctcaaat 180
Db 315 TTGGAATATCTCGATGTAGATGTAAGGACATGACAAATCTTAAACCAATTTCTCAA 374
Oy 181 aaactatgagctgtgtctctctctctctctctctctctctctctctctctctctctctct 240
Db 375 AAGTATAGGCTGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 434
Oy 241 gatatgagctgtgagagagagagagagagagagagagagagagagagagagagagagag 300
Db 435 GATATGACAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 494
Oy 301 attctctctctctctctctctctctctctctctctctctctctctctctctctctctct 360
Db 495 GTTTCAGTGGCTGAAAAAGTTAAACAAAGACTTGAATCTTTTCAGCAATGGAAGA 554
Oy 361 gatgagagagagagagagagagagagagagagagagagagagagagagagagagagagag 420
Db 555 GATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 614
Oy 421 gaaagagagagagagagagagagagagagagagagagagagagagagagagagagagag 480
Db 615 GGAGCATCGAGGAGCTGTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 674
Oy 481 ccaagctctctctctctctctctctctctctctctctctctctctctctctctctctctct 540
Db 675 CCAATGCTCTACCTCTGATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 734
Oy 541 gctcattatctcagaacacacacacacacacacacacacacacacacacacacacacacac 600
Db 735 GCTCTGTATTTTCCATGCTGATTTGATTAATAAGATCAGAGGTTTCTTAAGTTGATGG 794
Oy 601 aaaaatgaaagaaacacacacacacacacacacacacacacacacacacacacacacacac 660
Db 795 AAAAAATCAATGAAGAACCTCAGGATCTGAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 854
```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 19, 2002, 09:24:59 ; Search time 4158.4 Seconds
(without alignments)
2940.554 Million cell updates/sec

Title: US-09-763-292-2
Perfect score: 743
Sequence: 1 tcagaataatttgagcctg.....ttagctcatgtgaacgggg 743

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb.ba.*
- 2: gb.hig.*
- 3: gb.in.*
- 4: gb.om.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: qb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.hum.*
- 18: em.in.*
- 19: em.om.*
- 20: em.or.*
- 21: em.ov.*
- 22: em.pat.*
- 23: em.ph.*
- 24: em.pl.*
- 25: em.ro.*
- 26: em.sts.*
- 27: em.sy.*
- 28: em.un.*
- 29: em.vi.*
- 30: em.htgo_hum.*
- 31: em.htgo_inv.*
- 32: em.htgo_rod.*
- 33: em.htgo_hum.*
- 34: em.htgo_inv.*
- 35: em.htgo_rod.*
- 36: em.htgo_other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed. and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	739.8	99.6	743	9	HUM2C9X02
2	728.8	98.1	143087	2	LI6878 Homo sapien
3	728.8	98.1	173154	2	AL1359672 Homo sapi
4	635.6	85.5	205791	2	AL133513 Homo sapi
5	433.6	58.4	143087	2	AL583836 Homo sapi
6	346.4	46.6	156492	2	AL359672 Homo sapi
7	331.2	44.6	654	9	AL157835 Homo sapi
8	325	43.9	173154	2	HUM2C18X02
9	325	43.9	173154	2	AL133513 Homo sapi
10	280	37.7	711	4	AL583836 Homo sapi
11	264.6	35.6	463	9	M74201 Rabbit cyto
12	238.6	32.1	147838	2	AF136833 Homo sapi
13	225.8	30.4	400	11	AC038888 Rattus no
14	221.8	29.9	168457	2	GI4580 human STS S
15	217.2	29.2	513	10	AC078913 Mus muscu
16	178	24.0	163733	2	M18357 Rat cytochr
17	163	21.9	1473	22	AC083911 Rattus no
18	163	21.9	1473	22	E10631 Human CDNA
19	163	21.9	1576	9	E10853 CDNA encodi
20	163	21.9	1577	9	M15331 Human liver
21	163	21.9	1814	9	M21939 Human cytoc
22	163	21.9	1845	9	S46963 putative CY
23	163	21.9	1852	6	M61857 Human cytoc
24	163	21.9	1854	6	AR071579 Sequence
25	163	21.9	1854	9	AR071576 Sequence
26	161.4	21.7	1826	9	M61855 Human cytoc
27	161.4	21.7	1843	6	D00173 Homo sapien
28	153.4	20.6	1892	6	E02279 Human liver
29	152.4	20.5	1441	9	AR071581 Sequence
30	151	20.3	626	10	M21940 Human cytoc
31	149.6	20.1	1444	6	M33545 Rat female-
32	149.6	20.1	1473	22	EL4930 Artificial
33	149.6	20.1	1669	6	E10866 CDNA encodi
34	149.6	20.1	1746	6	E14931 Human mRNA
35	149.6	20.1	1746	9	AR071575 Sequence
36	149.6	20.1	2395	9	M61854 Human cytoc
37	144	19.4	200	6	L07093 Human cytoc
38	144	19.4	200	22	HUMSVPA450A
39	140.2	18.9	1473	22	AR048195 Sequence
40	140.2	18.9	1995	9	E10095 Synthetic m
41	140.2	18.9	2009	6	E10865 CDNA encodi
42	140.2	18.9	2258	6	M61856 Human cytoc
43	140.2	18.9	2258	9	AR071577 Sequence
44	135.8	18.3	1473	22	AR071580 Sequence
45	135.8	18.3	1473	22	M61853 Human cytoc
					E10639 Human CDNA
					E10862 CDNA encodi

ALIGNMENTS

RESULT 1	
HUM2C9X02	
LOCUS	HUM2C9X02 743 bp DNA PRI 08-FEB-1999
DEFINITION	Homo sapiens cytochrome P450C29 (CYP2C9) gene, exons 2 and 3.
ACCESSION	LI6878
VERSION	LI6878.1 GI:291608
KEYWORDS	CYP2C9; cytochrome P450; mephenytoin 4-hydroxylase.
SEGMENT	2 of 7
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE	1 (sites)
JOURNAL	Goldstein,J.A., Raucy,J.L., Blaisdell,J.A., Faletto,M.B. and
MEDLINE	Romkes,M.
REFERENCE	Cloning and expression of complementary DNAs for multiple members
	of the human cytochrome P4501C subfamily
	Biochemistry 30, 3247-3255 (1991)
	91182740
	2 (bases 1 to 743)

AUTHORS de Morais,S.M., Schweikl,H., Blaisdell,J. and Goldstein,J.A.
 TITLE Gene structure and upstream regulatory regions of human CYP2C9 and CYP2C18
 JOURNAL Biochem. Biophys. Res. Commun. 194 (1), 194-201 (1993)
 MEDLINE 93326116
 REFERENCE 3 (sites)
 AUTHORS Romkes,M., Faletto,M.B., Blaisdell,J.A., Raucy,J.L. and Goldstein,J.A.

TITLE Correction: Cloning and expression of complementary cDNAs for multiple members of the human cytochrome P4501C subfamily
 JOURNAL Biochemistry 32, 1390-1390 (1993)
 MEDLINE 93192243

FEATURES Location/Qualifiers
 source 1..743

exon /organism="Homo sapiens"

/db_xref="taxon:9606"

/tissue_type="liver"

/tissue_lib="EMBL3 library from J.A. Goldstein"

131..293

/gene="CYP2C9"

/citation=[1]

/evidence=experimental

465..614

/gene="CYP2C9"

/citation=[1]

/citation=[3]

/number=3

/evidence=experimental

BASE COUNT 182 a 152 c 196 g 213 t

ORIGIN

Query Match 99.6%; Score 739.8; DB 9; Length 743;
 Best Local Similarity 99.7%; Pred. No. P.1e-197;
 Matches 741; Conservative 0; Mismatches 2; Indels 0; Gaps 0:

OY 1 tcagaaatttgagcctgtgtgctgaataaaagcgtacatacaatacaataatca 60
 |||||
 DB 1 TCAGAAATATTTGAAGCCGTGTGCTGTAATAAAGCATACAAATACAAATATCA 60
 |||||
 OY 61 tgcataatcagccttagcaaaatggacaaatagtaacttcgttcttattctgtct 120
 |||||
 DB 61 TGTAAATCAGGCTTAGCAAAATGGACAAATAGTAACTTCGTTTCGTGTATCTCTGCT 120
 |||||
 OY 121 actttcctagctctcaaaagtctatggccctgtgttcactctgtattttggcctaaaacc 180
 |||||
 DB 121 ACTTCTAGCTCTCAAGGCTCTATGGCCCTGTGTTCACTCTGTATTTGGCCTCAAAACC 180
 |||||
 OY 181 catagtgtgtcgtatggtatgaagcagtggaaggaagccctgattgtcttgaggaga 240
 |||||
 DB 181 CATAGTGTGTCTGATGGATATGAAGCAGTGAAGGAAGCCCTGATTGATCTTGGAGAGGA 240
 |||||
 OY 241 gttttctggaagaggcattttccctctgctgctgaagagctaacacaggaatttggtagtg 300
 |||||
 DB 241 GTTTCTGGAGAGGCATTTTCCCACTGGCTCAAGAGCTAACAGAGATTTGGTAGGTG 300
 |||||
 OY 301 tgcattgctgtttcagcatctgtcttggttggttggttggttggttggttggttggttac 360
 |||||
 DB 301 TGCATGTGCTGTTTACAGCATCTGTCTGGGGATGGGAGGATGGAAGAGAGAGACTTAC 360
 |||||
 OY 361 agagctctctggcagagcttggtcccatccacatggtctccctcagcttctctcttt 420
 |||||
 DB 361 AGAGCTCTCGGCAGAGCTTGGCCCATCCACATGGCTGCCAGTGTCCAGTTCCTCTT 420
 |||||
 OY 421 cttgcttggtatctccctcctagtttcgtttctctctctctctctctctctctctctct 480
 |||||
 DB 421 CTTGCTTGGGATCTCCCTCTAGTTTCGTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 480
 |||||
 OY 481 tgaagaagaatgaagaagaatccggtttctccctctcatcagctgcgaattttgggat 540
 |||||
 DB 481 TGGAAAGAAATGGAGAGAGATCCGGGCTTCTCCCTCATGACGCTCGGAAATTTGGGAT 540
 |||||

OY 541 ggggaagaggagcattgagaccgtgttcaagagaagcccgctccttctgtgagagatt 600
 |||||
 DB 541 GGGGAAGAGAGCAGCAITGAGGACCGTGTTCAGAGGAAGCCCGCTCCTTGTGTGAGAGATT 600
 |||||
 OY 601 gagaaaaaaccaagggtgggtgacctactccatatacactgaccttacttgagactactatct 660
 |||||
 DB 601 GAGAAAAACCAAGGTTGGGTGACCTTACTCCATATACATGACCTTACTTGAGACTACTACT 660
 |||||
 OY 661 tctctactgacattcttgaaaacatttcagggtggccatatacttctcattatgagctcgg 720
 |||||
 DB 661 TCCTACTGACATTCITGGAACATTCAGGGGTGGCCATATCTTTTCATTATGAGTCTGG 720
 |||||
 OY 721 ttgttagctcatgtgaagcggg 743
 |||||
 DB 721 TTGTTAGCTCATGTGAGCGGG 743
 |||||

RESULT 2

AL359672

LOCUS

DEFINITION

PROGRESS ***, 2 unordered pieces.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

On Jun 28, 2001 this sequence version replaced gi:14529836.

----- Genome Center

Center: Sanger Centre

Web site: http://www.sanger.ac.uk

Contact: humquery@sanger.ac.uk

----- Project Information

Center project name: BA208C17

----- Summary Statistics

Assembly program: XGAP4; version 4.5

Sequencing vector: plasmid; L08752; 100% of reads

Chemistry: dye-terminator Big Dye; 100% of reads

Consensus quality: 142666 bases at least Q40

Consensus quality: 142843 bases at least Q30

Consensus quality: 142920 bases at least Q20

Insert size: 142987; sum-of-contigs

Insert size: 136404; agarose-fp

Quality coverage: 8.55x in Q20 bases; sum-of-contigs Quality

coverage: 7.82x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently

* consists of 2 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

* 1 47005: contig of 47005 bp in length

* 47006 47105: gap of 100 bp

* 47106 143087: contig of 95982 bp in length.

Location/Qualifiers

1..143087

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="10"

/clone="RP11-208C17"

/clone_lib="RPC1-11.1"

FEATURES

source

```
misc_feature 1..47005
/note="assembly_fragment:00494
clone_end:SP6
vector_side:left"
misc_feature 47106..143087
/note="assembly_fragment:00957"
BASE COUNT 41247 a 26155 c 27788 g 47797 t 100 others
ORIGIN
Query Match 98.1%; Score 728.8; DB 2; Length 143087;
Best Local Similarity 99.6%; Pred. No. 1.5e-194;
Matches 741; Conservative 0; Mismatches 2; Indels 1; Gaps 1;
Qy 1 tcagaaatattgaagcctgtgtgctgaataaaagcacaatacaatacaataatca 60
Db 60120 TCAGAAATATTGAAGCCTGTGTGCTGAATAAAGCATACAAATACAAATAICA 50179
Qy 61 tgcataatcagccttagcaaatgacaaatagtaacttcgtttgctgttatctctct 120
Db 60180 TGCTAAATCAGGCCTAGCAATGACAAATAGTAACCTCGTTTGCTGTAICTGCT 50239
Qy 121 acttcttagctctcaaggtctatgcccctgtgttcaactctgtattttggcctgaacc 180
Db 60240 ACTTCTTAGCTCTCAAGGCTATGCCCCTGTGTCACICTGTATTITGGCCCTGAACC 50299
Qy 181 catagtgtctcatagatgaagcagtaagcaagcagccctgattatctctgagaga 240
Db 60300 CATAGTGTGCTGATGATGAAGCAGTGAAGAACCCCTGATTGATCTTGAGAGGA 50359
Qy 241 gttttctggaagagcattttccacactggctgaaagagctaaacagagatttgtagtg 300
Db 60360 GTTTTCTGGAAGAGGCATTTTCCACCTGGCTGAAAGAGCTAACAGAGATTGGTAGTG 50419
Qy 301 tcatgtacctgtttcaagcatctcttgggatggagatggagatgaaacagaaacttac 360
Db 60420 TGCAATGCTGCTTTCAGCATCTGCTTGGGATGGGGAGGATGAAACAGACATAC 50479
Qy 361 agagctctcgggcagagcttggccatccacatggctgcccagtgccagcttctcttt 420
Db 60480 AGAGCTCTCGGCGAGAGCTTGGCCCATCCACATGGCTGCCAGTGTACAGTCTCTTT 50539
Qy 421 ctgctcaggatctccctcctagtttctctctctctctctctctctctctctctctct 480
Db 60540 CTGCTCGGATCTCCCTCCTAGTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 50599
Qy 481 tggaaagaaatggaagagatccgcgctttctccctcatgacgctgcggaattttggat 540
Db 60600 TGAAGAAGAAATGGAAGAGATCCGCGCTTCTCCCTCATGACGCTGCGGAATTTGGAT 50659
Qy 541 ggggaagagagcattgaggaccgtgttcaagaggaagccgcctctgtgaggaatt 500
Db 60660 GGGGAAGAGGAGCATTTGAGGACCGTGTTCAGAGAGGAGCCGCTGCTTGTGAGGAGTT 50719
Qy 601 gagaacacacaggtggtggtgacctactccatctactgacctactgagctactatct 560
Db 60720 GAGAAACCAAGGTTGGGTGACCTACTCTCAATACATGACCTTACTTGGACTACTATCT 50779
Qy 661 tctctactgacattcttggaaacatttcaggggtggccatatcttctcattatagagt-ctg 719
Db 60780 TCTCTACTGACATCTTGGAAACATTTTCAGGGGIGGCCATATCTTCTTATTATGAGTCTG 50839
Qy 720 gtttataactcatgaaagcagg 743
Db 60840 GTTGTAGCTCATGTGAACCGGGG 60863
RESULT 3
AL133513 173154 bp DNA HTG 04-JUL-2001
LOCUS Homo sapiens chromosome 10 clone RP11-400G3, *** SEQUENCING IN
DEFINITION PROGRESS ***, in ordered pieces.
ACCESSION AL133513
```

```
AL133513.11 G1:14575067
HTG: HTGS_PHASE2; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
human.
ORGANISM Homo sapiens
REFERENCE
AUTHORS Brown,J.
TITLE Direct Submission
JOURNAL Submitted (03-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Jun 28, 2001 this sequence version replaced gi:14456168.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA400G3
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 172989 bases at least Q40
Consensus quality: 173136 bases at least Q30
Consensus quality: 173145 bases at least Q20
Insert size: 173154; sum-of-contigs
Insert size: 116005; 33.1% error; agarose-fp
Quality coverage: 6.48x in Q20 bases; sum-of-contigs Quality
coverage: 9.80x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
source
1..173154
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP11-400G3"
/clone_lib="RPC1-11.2"
1..173154
/misc_feature
1..173154
/note="assembly_fragment:00748
vector_side:right"
BASE COUNT 49163 a 32272 c 33707 g 58012 t
ORIGIN
Query Match 98.1%; Score 728.8; DB 2; Length 173154;
Best Local Similarity 99.6%; Pred. No. 1.5e-194;
Matches 741; Conservative 0; Mismatches 2; Indels 1; Gaps 1;
Qy 1 tcagaaatattgaagcctgtgtgctgaataaaagcacaatacaatacaataatca 60
Db 150754 TCAGAAATATTGAAGCCTGTGTGCTGAATAAAGCATACAAATACAAATAICA 150813
Qy 61 tgcataatcagccttagcaaatgacaaatagtaacttcgtttgctgttatctctctct 120
Db 150814 TGCTAAATCAGGCCTTAGCAATGACAAATAGTAACCTCGTTTGCTGTAICTGCTCT 150873
Qy 121 acttcttagctctcaaggtctatgcccctgtgttcaactctgtattttggcctgaacc 180
Db 150874 ACTTCTTAGCTCTCAAGGCTATGCCCCTGTGTCACICTGTATTITGGCCCTGAACC 150933
Qy 181 catagtgtctcatagatgaagcagtaagcaagcagccctgattatctctgagaga 240
Db 150934 CATAGTGTGCTGATGATGAAGCAGTGAAGAACCCCTGATTGATCTTGGAGAGGA 150993
Qy 241 gtttctgaaagagcattttccacactggctgaaagagctaaacagagatttgtagtg 300
```

```

Db 150994 GTTTCTGGAAGAGGCAATTTCCCACTGGCTGAAAGAGCTAACAGAGGANTTTGGTAGGTG 151053
Qy 301 tgcattgctctgttccagcatctgtcttgggagatggggagatggaacacagagacttac 360
|||||
Db 151054 TGCATGTGCTGTCTTTCAGCATCTGTCTTGGGATGGGAGGATGGAAACACAGAGACTTAC 151113
|||||
Qy 361 agagctctctggcgaagattggcccatccacatggctccagtgctccctctctctt 420
|||||
Db 151114 AGAGCTCTCTGGGACAGAGTTGGCCCATCCACATGGCTGCCAGTGTCAGCTTCCCTTT 151173
|||||
Qy 421 ctgctctgggagatccctctcagttcgttctctctctctctctctctctctctctctcagcaa 480
|||||
Db 151174 CTGGCTGGGATCTCCCTCTAGTTCTGTTCTCTTCTCTTCTGTTAGGAATTTGTTTCAGCAA 151233
|||||
Qy 481 tgaagaagaaatggaagagatggcccatccacatggctccagtgctccctctctctctctctcagcaa 540
|||||
Db 151234 TGAAGAAGAAATGGAAGAGAGATCGGGGTTTCCCTCATGACGCTCGGAAATTTGGGAT 151293
|||||
Qy 541 ggggaagagagcattgagacagctgttcaagagagagccgctctctctctctctctctctctctctctct 600
|||||
Db 151294 GGGGAAGAGAGGAGGATGAGACCGTGTTCAGAGAGGAGGCGGCTGCTTCTGAGGAGT 151353
|||||
Qy 601 gagaaacacaaaggtgggtgacccctactccatcatcactgacacttactggaactactatct 660
|||||
Db 151354 GAGAAACACAAAGGTTGGGTGACCTTACTCCATATCAGTACGCTTACTGAGCTACTACT 151413
|||||
Qy 661 tctctactgacattctggaacatttcaagagtgagcattctctctctctctctctctctctctctctctct 719
|||||
Db 151414 TCTCTACTGACATCTTCTGGAACATTTGAGGGGTGGCCATATCTTTCATTAATGAGTCCCTG 151473
|||||
Qy 720 gttgttagctcatggaagcgggg 743
|||||
Db 151474 GTTGTAGTCATGTGAAGCGGGG 151497
|||||

```

```

RESULT 4
AL583836 205791 bp DNA HTG 13-JUN-2001
LOCUS Homo sapiens chromosome 10 clone RP11-466J14, *** SEQUENCING IN
PROGRESS ***, 2 unordered pieces.
ACCESSION AL583836
VERSION AL583836.11 GI:14455940
KEYWORDS HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Mashreghi-Mohammadi, M.
Direct Submission
Submitted (11-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerquest@sanger.ac.uk
On Jun 14, 2001 this sequence version replaced gi:14148873.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA466J14
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 205236 bases at least Q40
Consensus quality: 205448 bases at least Q30
Consensus quality: 205560 bases at least Q20
Insert size: 205691; sum-of-ctings
Insert size: 188405; 9.1% error; agarose-fp
Quality coverage: 11.31x in Q20 bases; sum-of-ctings Quality
Coverage: 12.57x in Q20 bases; agarose-fp
-----

```

* NOTE: This is a 'working draft' sequence. It currently consists of 2 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 88240: contig of 88240 bp in length
88241 88340: gap of 100 bp
88341 205791: contig of 117451 bp in length.

FEATURES
Location/Qualifiers
1..205791
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP11-466J14"
/clone_lib="RPC1-11.2"
1..88240
/note="assembly_fragment:00191"
fragment_chain:1
88341..205791
/note="assembly_fragment:01703"
fragment_chain:1
clone_end:T7
vector_side:right

BASE COUNT 55694 a 38262 c 40890 g 69845 t 100 others
ORIGIN

Query Match 85.5%; Score 635.6; DB 2; Length 205791;
Best Local Similarity 93.5%; Pred. No. 2.8e+168;
Matches 696; Conservative 0; Mismatches 44; Indels 4; Gaps 3;

Qy 1 tcagaaatattgagcctgtgtgctgaataaaagcatacaatacaatgaaatatca 60
|||||
Db 96936 TCAGAAATATTG-AGCCTGTGCTGAATAATAAGAGCATAACAATAAATAATGA 96994
|||||
Qy 61 tctcaaatcagccttagcaaaatggacaaatagtaactctgttctctctctctctct 120
|||||
Db 96995 ATCTAAGTCAGGCTTAGTAAATGGACAAACAGTACATCTTCTGTTAACTGTATCT 97054
|||||
Qy 121 acttctcagctctcaaaagctatggccctgtgttcactctgtattttggcctgaaacc 180
|||||
Db 97055 CCTTTCTAGCTCTCAAAAATCTATGGCCCTGTGTCTACTCTGTATTTGGCCTGGAACG 97114
|||||
Qy 181 catagtgtgtctcatggatataagacagatgaaggaagccctgattcttggaggaga 240
|||||
Db 97115 CATGTGGTGTCTGCATGGATATGAAGTGGTGAAGGAAGCCCTGATTTGAGAGGA 97174
|||||
Qy 241 gttttctgaaagagcattttccactggctggaagagctaacagaggtttggtagggt 300
|||||
Db 97175 GTTTCCTGGAAGAGGCCATTTCCCACTGCTGAAAGAGCTAACAGAGGATTTGGTAGGTG 97234
|||||
Qy 301 tgcattgctctgttccagcatctgtcttgggagatggggagatggaacacagagacttac 360
|||||
Db 97235 TGAAGTGGCTGTTTTCAGCATCTGTCTTGGGATGGGAGGATGGAAAC--AGACTAGC 97292
|||||
Qy 361 agagctctctggcgaagatcttgcccatccacatggctccagtgctccctctctctcagcaa 420
|||||
Db 97293 AGAGCTTCTCGGACAGAGCTTGGCCCATCCACATGGCTGCCAGTCTCAGCTTCTCTTT 97352
|||||
Qy 421 ctgctctgggagatccctctcagttctctctctctctctctctctctctctctctcagcaa 480
|||||
Db 97353 CTTGCTCTGGGATCTCCCTCTCCTAGTTTCGTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 97412
|||||
Qy 481 tgaagaagaaatggaagagatcccggtttctccctcatgacgctcgcggaattttgggat 540
|||||
Db 97413 TGAAGAAGATGGAAGAGAGATCGGGCTTTCTCCCTCATGACCTCGCGAAATTTGGGAT 97472
|||||
Qy 541 ggggaagagagcattgagagacagctgttcaagagagagcccgctgctctgtgagagatt 600
|||||
Db 97473 GGGGAAGAGGAGCATTCAGGACCGCTGTTCAGAGGAAGAGCCCGCTGCTTGTGGAGGAGTT 97532
|||||


```

ORGANISM      Homo sapiens
               Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
               1 (bases 1 to 156492)
               Blakey,S.
REFERENCE      Submitted (05-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
AUTHORS        CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
TITLE          requests: clonerquest@sanger.ac.uk
JOURNAL        On Jul 8, 2001 this sequence version replaced gi:14585042.

COMMENT       ----- Genome Center
               Center: Sanger Centre
               Center code: SC
               Web site: http://www.sanger.ac.uk
               Contact: humquery@sanger.ac.uk
               ----- Project Information
               Center project name: BA361K9
               ----- Summary Statistics
               Assembly program: XGAP4; version 4.5
               Sequencing vector: plasmid; L08752; 100% of reads
               Chemistry: Dye-terminator Big Dye; 100% of reads
               Consensus quality: 156104 bases at least Q40
               Consensus quality: 156214 bases at least Q30
               Consensus quality: 156256 bases at least Q20
               Insert size: 156492; sum-of-contigs
               Insert size: 150975; 13.7% error; agarose-fp
               Quality coverage: 8.34x in Q20 bases; sum-of-contigs Quality
               coverage: 9.69x in Q20 bases; agarose-fp
               -----
               * NOTE: This is a 'working draft' sequence.
               * This sequence will be replaced
               * by the finished sequence as soon as it is available and
               * the accession number will be preserved.
               Location/Qualifiers
                 1..156492
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /chromosome="10"
                    /clone="Rp11-361K9"
                    /clone_lib="RPC1-11.2"
                 1..156492
                    /note="assembly_fragment:03685
                    clone_end:SP6
                    vector_side:left"
EASE COUNT    43071 a 30007 c 31050 g 52364 t
CIGIN

Query Match   46.6%; Score 346.4; DB 2: Length 156492;
Best Local Similarity 78.1%; Pred. No. 9.7e-87;
Matches 478; Conservative 0; Mismatches 106; Indels 28; Gaps 4

Oy 38 atacaatacaatgaaataatcatgtaataatcaggcttagcaaaatggacaaatagtaac 97
Db 103686 ACAACATATGATGATATATATGCTGTAATGACGACCAATATATAGCAAACTACTGAG 103745

Oy 98 ttcgtttgctgtt---atctctgtctactttccctagctctcaagggtcatggccctgtg 154
Db 103746 TTTTGTACTATTGTGAACCTCCTTTCTATGTTTACTTCTCAAAAGCTATGCGCCCTGTG 103805

Oy 155 ttcactctgtatttggcctgaaacccatagtgctctcatgatatgaagcagtgtaag 214
Db 103806 TTCACGTGTGATTTTGGCTCGAAGCCCATTTGTGGTGTTCATGTGGATATGAACGACGTGAAG 103865

Oy 215 gaagccctgattgatcttgaggagagattttcttggaaaggacattttcccactgctgaa 274
Db 103866 GAGGCCCTGATTGTATCATGATGGAGAGGAGCTTTTCTGGAAGAGGAAGTTTCCAGTGGCTGAA 103925

Oy 275 agagctaacagagatttggttaggtgtgcattgacctgtt----- 314
Db 103926 AAGGTTAACAAGAGGACATTGGTAAATGTTTGGTATGTCATGTATCTGTGTGTAATGTTACATGTGTAATG 103985

Oy 315 --tcagcatctctctgtggagtgaggagatggaacacagagacttacagagctctcgg 372

```



```
Db 419 CAGGCTTCATCGCATCTGCTGTCAGCCC 451
|| |||| | | ||||| ||| |
AC083888 147838 bp DNA HTG 01-APR-2001
Rattus norvegicus clone RP32-307N15, WORKING DRAFT SEQUENCE, 21
unordered pieces.
AC083888
AC083888.12 GI:13493083
HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE
Norway rat
ORGANISM
Rattus norvegicus
REFERENCE
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
(bases 1 to 147838)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,
Alsbrooks,S.L., Anaratunga,H.C., Are,J.R., Banks,I., Barbarella,J.,
Benton,J., Binage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,I.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,I.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hoques,M., Holloway,C.,
Hollins,B., Honsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Loulsegh,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M.,
Mei,G., Metker,M., Miner,G., Miner,Z., Mitchell,I., Mohabbat,K.,
Morgan,A., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokwenkwo,S.,
Ogulu,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M.,
Ruiz,S., Savary,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
Tang,H., Tansey,J., Taylor,C., Taylor,F., Telford,B., Thomas,N.,
Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
Washington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 147838)
Worley,K.C.
Direct Submission
Submitted (05-OCT-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Apr 1, 2001 this sequence version replaced gi:13253418.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: IUAF
FEATURES
Source
1..147838
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/clone="RP32-307N15"
BASE COUNT 4949 a 29290 c 29368 g 42091 t 2140 others
ORIGIN
1
22638: contig of 22638 bp in length
22739: gap of unknown length
22739: contig of 16838 bp in length
39576: gap of unknown length
39576: gap of unknown length
59439: contig of 19763 bp in length
59439: gap of unknown length
59539: gap of unknown length
78333: contig of 18694 bp in length
78333: gap of unknown length
87924: contig of 9591 bp in length
87924: gap of unknown length
88024: gap of unknown length
88025: contig of 10015 bp in length
88025: gap of unknown length
98140: contig of 10987 bp in length
109126: gap of unknown length
109127: gap of unknown length
109227: contig of 6387 bp in length
115614: contig of unknown length
115614: gap of unknown length
121553: contig of 5840 bp in length
121553: gap of unknown length
121554: gap of unknown length
121554: contig of 4132 bp in length
125785: gap of unknown length
125786: gap of unknown length
125885: gap of unknown length
130841: contig of 4756 bp in length
130841: gap of unknown length
133125: contig of 2384 bp in length
133125: gap of unknown length
133225: gap of unknown length
133225: contig of 2201 bp in length
135427: gap of unknown length
135427: contig of 2036 bp in length
137527: gap of unknown length
137527: contig of 1613 bp in length
137563: gap of unknown length
139375: gap of unknown length
139375: contig of 1729 bp in length
141105: gap of unknown length
141105: contig of unknown length
141205: contig of 1286 bp in length
142491: gap of unknown length
142491: contig of 1466 bp in length
144057: gap of unknown length
144057: contig of 1316 bp in length
144157: gap of unknown length
144157: contig of unknown length
145473: gap of unknown length
145473: contig of 1095 bp in length
146573: gap of unknown length
146573: contig of unknown length
146768: contig of 1071 bp in length.
146768: Location/Qualifiers
1..147838
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/clone="RP32-307N15"
Center clone name: RP32-307N15
----- Summary Statistics
Sequencing vector: Plasmid: M77789
Chemistry: Dye-terminator Big Dye: 89% of reads
Assembly program: Phrap: version 0.990329
Consensus quality: 133494 bases at least Q40
Consensus quality: 142576 bases at least Q30
Consensus quality: 147441 bases at least Q20
Estimated insert size: 136695; sum-of-contrigs estimation
Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 4.7x in Q20 bases; sum-of-contrigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 21 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1
22638: contig of 22638 bp in length
22739: gap of unknown length
22739: contig of 16838 bp in length
39576: gap of unknown length
39576: gap of unknown length
59439: contig of 19763 bp in length
59439: gap of unknown length
59539: gap of unknown length
78333: contig of 18694 bp in length
78333: gap of unknown length
87924: contig of 9591 bp in length
87924: gap of unknown length
88024: gap of unknown length
88025: contig of 10015 bp in length
88025: gap of unknown length
98140: contig of 10987 bp in length
109126: gap of unknown length
109127: gap of unknown length
109227: contig of 6387 bp in length
115614: contig of unknown length
115614: gap of unknown length
121553: contig of 5840 bp in length
121553: gap of unknown length
121554: gap of unknown length
121554: contig of 4132 bp in length
125785: gap of unknown length
125786: gap of unknown length
125885: gap of unknown length
130841: contig of 4756 bp in length
130841: gap of unknown length
133125: contig of 2384 bp in length
133125: gap of unknown length
133225: gap of unknown length
133225: contig of 2201 bp in length
135427: gap of unknown length
135427: contig of 2036 bp in length
137527: gap of unknown length
137527: contig of 1613 bp in length
137563: gap of unknown length
139375: gap of unknown length
139375: contig of 1729 bp in length
141105: gap of unknown length
141105: contig of unknown length
141205: contig of 1286 bp in length
142491: gap of unknown length
142491: contig of 1466 bp in length
144057: gap of unknown length
144057: contig of 1316 bp in length
144157: gap of unknown length
144157: contig of unknown length
145473: gap of unknown length
145473: contig of 1095 bp in length
146573: gap of unknown length
146573: contig of unknown length
146768: contig of 1071 bp in length.
146768: Location/Qualifiers
1..147838
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/clone="RP32-307N15"
BASE COUNT 4949 a 29290 c 29368 g 42091 t 2140 others
ORIGIN
```


Db	27819	TTTGCTTAATAAAAAATAAAATACAGACATAATATCAGCATGAATGAGGTTACACT	27878
Qy	81	atgagcaaaatagtaac-ttcgtttgtgttatctctgtctacttctct-----agctc	133
Db	27879	ATAATTGATCAAGGTTACATAATGTTGCTAATAATTTGCAATTTCTTATTTCCAGTTC	27938
Qy	134	tcaaaaggtctatggccctgtgttcactctgtatcttatttggcctgaaacccatagtggtctg	193
Db	27939	TCAAAAGCTTATGGCCGTGTGTTCACTCTGTATTGGGCTCAAAAGCCTACTGTGATATTG	27998
Qy	194	catgatataagcagcagtgaaagaaacccctgtatgtatcttggagagagatttcttggaaa	253
Db	27999	CATGGGTATGAAGCAGTGAAGGAAGCTCTGATTGATCGTGGGAGGAGTTTGCTGGTAGA	28058
Qy	254	ggcattttccactggctgaaagagctaacagagagatttgtagtggtgcattgacctgt	313
Db	28059	GGAGCTTTCCAAATGGCTGCAAAAATTTATTAAGGTTTGTGTAAGT----ATGTTACAT	28114
Qy	314	ttcagcatctgtcttgggagtgaggagatggaacacagaaacttacagagctctctcgg	373
Db	28115	ATTATGTGTGTTTCACAGTGTTTACAGTGAGATATAAGATTGAAAGAAAGGATGGG	28174
Qy	374	cagagcttgagccatccatagcgtgccagtgctcagcttctcttcttctgctgggac	433
Db	28175	CTTCTCTGTGAGGCAATGGGCCACTCTATGTGTTTCCACACTCACCTGTCTTAGCAT	28234
Qy	434	tcctctctagtttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt	493
Db	28235	CTATTCTAGTACC--TTCTCTTTCTTTAGCGTGTGTTTGTAGCAATGGAACAGATGG	28292
Qy	494	aaggagatccggcgttcttcttcttcttcttcttcttcttcttcttcttcttcttctt	553
Db	28293	AAAGATATGAGAGATTTACACTCATGACCCCTGCGAAATTTGGGATGGGAAAGGAAC	28352
Qy	554	attgagaccgtgttcaagaggagccgctgcttcttcttcttcttcttcttcttcttctt	613
Db	28353	ATTGAGACCGTGTTCAGAGGAAGACACAGTGTCTTGTGGAAAGACTGAGGAAACCAAA	28412
Qy	614	ggt 616	
Db	28413	GGT 28415	
RESULT 15			
RATCY45M2			
LOCUS			
Rat cytochrome P-450(M-1) gene, exons 2 and 3.			
ACCESSION			
M18357			
VERSION			
M18357.1 GI:203687			
KEYWORDS			
cytochrome P450.			
SEGMENT			
2 of 8			
SOURCE			
Rat (male) liver DNA, (library of L.Jagodzinsky), clones			
lambda-[1,3,18,22,23,29].			
ORGANISM			
Rattus norvegicus			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;			
Rattus.			
REFERENCE			
AUTHORS			
Morishima,N., Yoshioka,H., Higashi,Y., Sogawa,K. and			
Fujii-Kuriyama,Y.			
TITLE			
Gene structure of cytochrome P-450(M-1) specifically expressed in			
male rat liver			
JOURNAL			
Biochemistry 26, 8279-8285 (1987)			
MEDLINE			
88163490			
FEATURES			
source			
1..513			
/organism="Rattus norvegicus"			
/db_xref="taxon:10116"			
intron			
<1..11			
prim_transcript			
<1..>513			
/note="P450 mRNA and introns"			
exon			
12..1174			

Intron		/number=2	
		175..318	
exon		/note="P450 intron B"	
		319..468	
Intron		/number=3	
		469..513	
BASE COUNT		119 a 95 c 144 g 154 t	
ORIGIN		About 1.7 kb after segment 1.	
Query Match 29.2%; Score 217.2; DB 10; Length 513;			
Best Local Similarity 69.9%; Pred No 2.4e-50;			
Matches 348; Conservative 0; Mismatches 123; Indels 27; Gaps 3;			
Qy	121	acttctcagctctcaaaagctctatggccctgtgctcactgtgttccactctgtatcttggcctgaaacc	180
Db	2	AATTTTCCAGTTTTTCAAAAGCTACGGCCCTATATTACTCTGTATTGGGCTGAAGCC	61
Qy	181	catagtgtgctcagatgatataagcagtgaaagagccctgattgattcttggagaga	240
Db	62	CTTTGTGGTGTTCATGGGTATGAAGCTGTGAAGGAAGCTCTTGTGTATCTAGGAGAGA	121
Qy	241	gtttctggaagagcattttccctcactggctgaaagagcttaacacagagatttggtaggtg	300
Db	122	ALTTCTGGAAGAGCGAGTTTCCAGTATCTGAAAGAGTTTAAACAAGGCCCTTGGTAAGTG	181
Qy	301	tgcatgtccctgttctcagcatctgtcttggggatgggagagaiggaacacagacttacc	360
Db	182	TGAGTACAAATGTAATTTCTGTGTGTGTGAGCCCTCGCCCTTGGCAGA----TGAGACTTAA	237
Qy	361	agagctctcgggcagagcttgcccatccacatggctgccagtgctcagcttctctctt	420
Db	238	TGAGCTCAGAGCCATTAGGATAGC---ACCATGTGTGAGTTCGTGCTCGCTCCTCTGACT	294
Qy	421	cttgcttgggagctccctcctcctagtttcttcttcttcttcttcttcttcttcttcttctt	480
Db	295	AGT-----TTCTGTTTCCATTCTGGTAGGAGTCAITTTTAGCAA	334
Qy	481	tggaagaaatggagagatccagcgtttctcctctcatgacgtgcggaattttgggat	540
Db	335	TGGGATGCAATGGAAGGAGATCCGGCGTTTCTCCATCATGACCCCTGAGGACTTTTGGGAT	394
Qy	541	ggggaagagagcattttaggacctgttcaagagagagcccgctgaccttggagagatt	600
Db	395	GGGCAAGAGGACCATTCAGGACCGTATTCAAGAGGAGGAGCTCAGTGCCTTGTGGAGGAAC	454
Qy	601	gagaaaaaaccaaggtgg	618
Db	455	GAGGAAGAACCAAGGTAG	472

Search completed: April 19, 2002, 09:32:08
Job time: 7828 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 19, 2002, 08:08:00 ; Search time 232.27 Seconds
(without alignments)
885.492 Million cell updates/sec

Title: US-09-763-292-1
Perfect score: 1745
Sequence: 1 cttcaatggtacctttgtg.....acatattattattaaataga 1745

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 78474 seqs, 58898373 residues

Total number of hits satisfying chosen parameters: 156948

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending Patents_NA_New:*

- 1: /cgn2_6/ptodata/2/pna/PCT_NEW_COMB.seq:*
- 2: /cgn2_6/ptodata/2/pna/US06_NEW_COMB.seq:*
- 3: /cgn2_6/ptodata/2/pna/US07_NEW_COMB.seq:*
- 4: /cgn2_6/ptodata/2/pna/US08_NEW_COMB.seq:*
- 5: /cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq:*
- 6: /cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq:*
- 7: /cgn2_6/ptodata/2/pna/US60_NEW_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	139.6	8.0	526	US-10-106-698-2440	Sequence 2440, Ap
2	137.2	7.9	591	US-10-106-698-858	Sequence 858, App
3	59.2	3.4	465	US-10-103-313-627	Sequence 627, App
4	51.8	3.0	1186	US-10-103-313-621	Sequence 621, App
5	51.8	3.0	1186	US-10-105-299-7204	Sequence 7204, Ap
6	51.6	3.0	1454	US-09-559-013D-11	Sequence 11, Appl
7	51.6	3.0	1464	US-09-559-013-11	Sequence 11, Appl
8	49	2.8	2026	US-10-106-698-368	Sequence 368, App
9	49	2.8	13327	US-10-105-299-7201	Sequence 7201, Ap
10	45.8	2.6	1530	PCT-US02-06912-12	Sequence 12, Appl
11	45.8	2.6	1693	PCT-US02-06912-8	Sequence 8, Appl
12	44.6	2.6	254	US-09-975-254-18019	Sequence 18019, A
13	43.6	2.5	832	US-10-103-313-626	Sequence 626, App
14	43.6	2.5	1001	US-10-103-313-625	Sequence 625, App
15	41.4	2.4	272	US-09-975-254-16494	Sequence 16494, A
16	41.4	2.4	1182	US-10-103-313-622	Sequence 622, App
17	41.4	2.4	1182	US-10-105-299-7203	Sequence 7203, Ap
18	41	2.3	583	US-10-102-806-124	Sequence 124, App
19	41	2.3	602	US-10-103-313-66	Sequence 6, Appl
20	41	2.3	1745	PCT-US02-06912-6	Sequence 11513, A
21	40.8	2.3	252	US-09-975-254-11513	Sequence 252, App
22	40.8	2.3	248	US-09-975-254-21714	Sequence 21714, A
23	40	2.3	252	US-09-975-254-23845	Sequence 23845, A
24	39.4	2.3	237	US-09-975-254-14003	Sequence 14003, A
25	39.4	2.3	2355	US-10-103-313-103	Sequence 103, App
26	38.4	2.2	266	US-09-975-254-20507	Sequence 20507, A

ALIGNMENTS

RESULT 1

US-10-105-698-2440
Sequence 2440, Application US/10106698

GENERAL INFORMATION:

APPLICANT: Ruben et al.
TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptides
FILE REFERENCE: PA005PI
CURRENT APPLICATION NUMBER: US/10/106,698
CURRENT FILING DATE: 2002-03-27
PRIOR APPLICATION NUMBER: PCT/US00/26524
PRIOR FILING DATE: 2000-09-28
PRIOR APPLICATION NUMBER: US 60/157,137
PRIOR FILING DATE: 1999-09-29
PRIOR APPLICATION NUMBER: US 60/163,280
PRIOR FILING DATE: 1999-11-03
NUMBER OF SEQ ID NOS: 8564
SOFTWARE: PatentIn Ver. 3.0
SEQ ID NO 2440
LENGTH: 526
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (4)..(4)
OTHER INFORMATION: n equals a.t.g, or c
NAME/KEY: misc_feature
LOCATION: (23)..(23)
OTHER INFORMATION: n equals a.t.g, or c
NAME/KEY: misc_feature
LOCATION: (32)..(33)
OTHER INFORMATION: n equals a.t.g, or c
NAME/KEY: misc_feature
LOCATION: (310)..(310)
OTHER INFORMATION: n equals a.t.g, or c
NAME/KEY: misc_feature
LOCATION: (332)..(332)
OTHER INFORMATION: n equals a.t.g, or c
NAME/KEY: misc_feature
LOCATION: (336)..(336)
OTHER INFORMATION: n equals a.t.g, or c
NAME/KEY: misc_feature
LOCATION: (369)..(369)
OTHER INFORMATION: n equals a.t.g, or c
NAME/KEY: misc_feature
LOCATION: (375)..(375)
OTHER INFORMATION: n equals a.t.g, or c
NAME/KEY: misc_feature
LOCATION: (387)..(387)
OTHER INFORMATION: n equals a.t.g, or c

Sequence 1, Appl
Sequence 1380, A
Sequence 19229, A
Sequence 23331, A
Sequence 19290, A
Sequence 102, App
Sequence 25520, A
Sequence 7189, Ap
Sequence 1, Appl
Sequence 16, Appl
Sequence 12971, A
Sequence 4, Appl
Sequence 100, App
Sequence 24506, A
Sequence 202, App
Sequence 25, Appl
Sequence 9415, Ap
Sequence 3502, Ap
Sequence 1347, Ap

27 38 2.2 1756 6 US-10-104-706-1
28 37.5 2.2 253 5 US-09-975-254-13380
29 37.2 2.1 261 5 US-09-975-254-19229
30 36.5 2.1 286 5 US-09-975-254-23331
31 36.2 2.1 267 5 US-09-975-254-19290
32 35.8 2.1 745 6 US-10-103-313-102
33 34.2 2.0 191 5 US-09-975-254-25520
34 34.2 2.0 7932 6 US-10-105-299-7189
35 33.8 1.9 1658 6 US-10-026-666-1
36 33 1.9 2650 1 PCT-US02-08992-16
37 32.8 1.9 257 5 US-09-975-254-12971
38 32.8 1.9 1614 1 PCT-US02-06912-4
39 32.4 1.9 3220 6 US-10-103-313-100
40 32.2 1.8 265 5 US-09-975-254-24906
41 32 1.8 1084 6 US-10-103-313-202
42 32 1.8 5115 5 US-09-630-630B-25
43 31.8 1.8 4540 6 US-10-105-299-9415
44 31.6 1.8 358 6 US-10-106-698-3502
45 31.6 1.8 748 6 US-10-106-698-1347

US-10-103-313-627

Query Match 3.4%; Score 59.2; DB 6; Length 465;
Best Local Similarity 47.6%; Pred. No. 2.5e-06;
Matches 175; Conservative 0; Mismatches 193; Indels 0; Gaps 0;

Oy 468 gagtgagaaaaaaccaaggcttaccctgtgacccacttctcatctcctgggtgtgtctccc 527
Db 10 gaaatgcaaaagcaggaagaccctcttgccttcttccatcatcagcaatgccgtc 69
Oy 528 tgaatgtgactctcattatttccagaaacgcttccatataaagatcacgaattt 587
Db 70 tctaacatcttctcctgtctgtcttggccagcgttttgattacataatagtgaattc 129
Oy 588 cttaactgtgaaaaaattgaaataacatcaggatttgaagcaccctggatccag 647
Db 130 aagaaatgtgttttattgtcagaggctagaatactgtctgaacagtcagtcctc 189
Oy 648 atatgcaataatttcccaactatcattgattatttccggagaaccatacaaatattc 707
Db 190 ctggtcaacatctcccttggctttattaccttcccttggaccatttaaggaattaga 249
Oy 708 aaaaacctgtctttatggaagtgtattttggagaaaglaaaagaaacacccaagaatcg 767
Db 250 caaattgaaaaggatataaacacagtttctttaaataaatcatcaaaagaccatcaagtgct 309
Oy 768 atggacatcaaacacctcggacttatttattgtattgtcttcttgatcaaaaatgagaagaa 827
Db 310 ctggatagagaaacctcaggacttcatagacatgtaccttctccacatggaagagag 369
Oy 828 aagcaaaa 835
Db 370 aggaaaaa 377

RESULT 4
US-10-103-313-621
: Sequence 621, Application US/10103313
: GENERAL INFORMATION:
: APPLICANT: Rosen et al.
: TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
: FILE REFERENCE: PJ207C1
: CURRENT APPLICATION NUMBER: US/10/103.313
: CURRENT FILING DATE: 2002-03-12
: NUMBER OF SEQ ID NOS: 653
: Prior Application removed - See File Wrapper or Palm
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 621
: LENGTH: 1186
: TYPE: DNA
: ORGANISM: Homo sapiens
US-10-103-313-621

Query Match 3.0%; Score 51.8; DB 6; Length 1186;
Best Local Similarity 55.2%; Pred. No. 0.0003;
Matches 101; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

Oy 977 ggaagagattgaactgttcattggcagaaacggagccctgcacgcacagcagggcca 1036
Db 5 qcaggagctggaccctgtgttaggtggaggccgcgcacacagcctggactatcgctgtg 64
Oy 1037 catgcctacacagatgctgtgtgtgacagaggttcacagagatacatcgacctatccccac 1096
Db 65 cctgcctacgccaacgagtgctgtcgtcagatccagtcagtcctcatcagcgtgggtgccct 124
Oy 1097 cagctgcctcctcagtcagtcagtcgtgagctttaaattcagaactacacctattccccagg 1156
Db 125 ggggtgcgcgcacacctcaccctgcagacaccacactgcacagcactgtctgtcccaagg 184
Oy 1157 cac 1159
||

Db 185 tac 187

RESULT 5
US-10-105-299-7204/c
: Sequence 7204, Application US/10105299
: GENERAL INFORMATION:
: APPLICANT: Rosen, et. al
: TITLE OF INVENTION: Human Secreted Proteins
: FILE REFERENCE: PS950
: CURRENT APPLICATION NUMBER: US/10/105,299
: CURRENT FILING DATE: 2002-03-26
: NUMBER OF SEQ ID NOS: 15197
: Prior Application removed - See File Wrapper or Palm
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 7204
: LENGTH: 1186
: TYPE: DNA
: ORGANISM: Homo sapiens
US-10-105-299-7204

Query Match 3.0%; Score 51.8; DB 6; Length 1186;
Best Local Similarity 55.2%; Pred. No. 0.0003;
Matches 101; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

Oy 977 ggaagagattgaactgttcattggcagaaacggagccctgcacgcacagcagggcca 1036
Db 1182 GCAGGACTGGACCCCTGTGTAGGTGGAGGCGCGCCCAAGGCTGGACTATCGCTGTG 1123
Oy 1037 catgcctacacagatgctgtgcacagaggttcacagagatacatcgacctatccccac 1096
Db 1122 CTGCTCTACGCCAACGCAGTGTCTCGAGATCCAGTCTTCATCAGCGTGTGCTCCCT 1063
Oy 1097 cagctgcctcctcagtcagtcagtcgttaattcagaactacacctattccccagg 1156
Db 1052 GGGGCTGGCGGACCCCTCACCTCGACACCCACCTGTCAGGCTGTCTGCCCAAGG 1003
Oy 1157 cac 1159
||

RESULT 6

US-09-559-013D-11/c
: Sequence 11, Application US/09559013D
: GENERAL INFORMATION:
: APPLICANT: Ono, Toshiro
: APPLICANT: Nakayama, Eiichi
: TITLE OF INVENTION: CANCER ASSOCIATED ANTIGENS AND USES
: FILE REFERENCE: L0461/7086
: CURRENT APPLICATION NUMBER: US/09/559,013D
: CURRENT FILING DATE: 2000-04-26
: PRIOR APPLICATION NUMBER: US 60/168,353
: PRIOR FILING DATE: 1999-12-01
: NUMBER OF SEQ ID NOS: 32
: SOFTWARE: FastSEQ for Windows Version 3.0
: SEQ ID NO 11
: LENGTH: 1454
: TYPE: DNA
: ORGANISM: Mus musculus
US-09-559-013D-11

Query Match 3.0%; Score 51.6; DB 5; Length 1454;
Best Local Similarity 44.0%; Pred. No. 0.00037;
Matches 329; Conservative 0; Mismatches 404; Indels 15; Gaps 2;

Oy 542 atccagatatgcataatttccacactatcttgatttcccggaacccatacaaaa 701
Db 854 ATCCAATCTTGACGATANTTCCCTTCTCAGGTTCTTCCCAACCCAGGCTCCAGAG 805

Qy	702	ttaacttaaaaccttgccttttatggaagtgatattttggagaagttaaagaacaccaca	761
Db	804	CTGAAGCAGATCCAAAGAGAGTCGGGACCATTATGTCAACGACGACGTGAACGGCGCACAG	745
Qy	762	gaatcatgggatcaatacaaacacctcggaactttaattgatgttcctcctgatacaaaatggag	821
Db	744	GACAGCCTGGTTTGCAGGCCAATGGAAAGACATGATTGACTACATCTCTCCAGGGAGTGCG	695
Qy	822	aaggaaaaaccccccaacagtccta----attcacatttgaaaacttgggttaatacactgcca	878
Db	684	AAGCAAAGGATGGCAAGACGAAGAGCGGTCTCCACGAGGGGACGTGCACATGTGCGTG	525
Qy	879	gctgacttacttggagctgggacagagacaacaacccctcgagatatgtctctcctt	938
Db	624	GTGGACCTGTTCAICGGCGGACCCGAGACCGGCTACCCAGCTCTCCTGGGCTGTGGCT	565
Qy	939	ctcctgctaaagcccacccagaggtcacagctaaagtccagggaagagattgaacgtgtcatt	998
Db	564	TTCCTGCTTCACCCACCTTGAGATCCAGAAGGACACTGCAGGAAGAGTTAGACCTCAAGCTG	505
Qy	999	ggcagaaaccggagcccttgcatgcacacaggggccacatccctcacagatagctgtg	1058
Db	504	GGCCCAGGCTCCACGCTCCTGTACAGGAACCCGATGTCAGCTGCCCTTGTCTCATGGCCACC	445
Qy	1059	gtcacagaggtccagagatcacatcgacctcatccccacacagcctgcccctcagagtgaacc	1118
Db	444	ATTGCCGAGGTGCTGGCTTTGGCGGCTGTGGTGCCCTTGGCCCTTGCCCCATCGTGCAACT	385
Qy	1119	tgtgacgttaaatctgaaacttacctcattcccgaaggcacaacatatataacttccctc	1178
Db	384	AGGGCTAGACGATCTCCGGCTATGACATCCCTAAGGATATGGTCAITCATCCCCAACATC	325
Qy	1179	actctgtgctacatgacacaacaaagaatttcccacccagagatgttgacctctgtcac	1238
Db	324	CAAGGCGCCAACCTTGEATCAGATGGTTTTGGGAACCTGCCACGAAGTTCTGSCCAGATCGC	265
Qy	1239	ttcttgatgaagtggaattttaagaaaaagtaactactcttcagctttctcagcagga	1298
Db	264	TTCTTGGAACCTTGGGAAGAAATCCCAGAACAACATCCTTTGGCTG-----TGGG	217
Qy	1299	aaacggatttgtggagagggcctggcccgcatggagctgtttttattctcagacttc	1358
Db	216	GCACGGCTGTGCTGSGAGAGCCTCTGSCAGGGCTGGAGCTCTTGTGCTGCTGGCTCGT	157
Qy	1359	attttacagaactttaacctgaattctc	1386
Db	156	CTGCTGACGGCCTTCACTCTGTGCTC	129

RESULT 7
US-09-559-013-11/c

US 03 255 013 11/03
; Sequence 11, Application US/09559013
; GENERAL INFORMATION:

APPLICANT: Ono, Toshio

APPLICANT: Nakayama, Eiichi

; TITLE OF INVENTION: CANCER ASSOCIATED ANTIGENS AND USES

; TITLE OF INVENTION: THEREFOR

; FILE REFERENCE: L0461/7086

; CURRENT APPLICATION NUMBER: US/09/559,013

; CURRENT FILING DATE: 2000-04-26

; PRIOR APPLICATION NUMBER: US 60/168,353

; PRIOR FILING DATE: 1999-12-01

; NUMBER OF SEQ ID NOS: 32

```

; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 11

```

```

: SEQ ID NO 11
: LENGTH: 146

```

```

: LENGTH: 1464
: TYPE: DNA

```

TYPE: DNA
ORGANISM: M...

ORGANISM: MUS MUSCULUS
MS-09-559-013-11

IT-610-6CC-60-60

```
Query Match          3.0%; Score 51.6; DB 5; Length 1464;
Best Local Similarity 44.0%; Pred. No. 0.00037;
```

	Matches	329;	Conservative	0;	Mismatches	404;	Indels	15;	Gaps	2;
Qy	642	atccagatatcaataattttcccactatcatcattgattatttcccgggaacccataacaaa	701							
Db	954	ATCCAAATCTTGACGATAATTCCTCTCAGGTTCTCTCCCAACCCAGGCTCCAGAAG	805							
Qy	702	ttaactaaaaaccttgcttttatggaagtgatattttggagaagtaaaagaacaccaa	751							
Db	804	CTGAAGCAGATTCCAAGAGAGTGGGAGCCATTATTGTCAAGCAGCAGCTGAAGCGGCACAA	745							
Qy	762	gaatcgtatggacatcaacaacccctcgggaacttattgattgcttccctgatcaaatgag	821							
Db	744	GACAGCTGGTTGCAGGCCAATGAAAGACATGATTGACTACATGCTCCAGGGAGTGGAG	685							
Qy	822	aagaaagcaaaaccaaagctctga---attcactattgaaaacttggtaactcactgca	878							
Db	584	AAGCAAGGGGATGGCAAAAGACAAGAGCGGCTCCACGAGGGGCACGTGCACATGTCGGTG	625							
Qy	879	gctgaacttaacttgagactggagagagagacaacagcacacaacctgagatatgctctcctt	938							
Db	624	GTGAGCTGTTTCATCGGCGGCACCGAGACCAGCGCTACACAGCTCTCTCTGGGCTGTGGCT	565							
Qy	939	ctcctgctgaagcacccagaggtcacagctaaagtccaggaagagattgaacgtgctatt	998							
Db	564	TTCTGCTTACCACCTTGAGATCCAGAAGCGCACTGCAGGAAGAGTTAGACCTCAAGCTG	505							
Qy	999	ggcagaacccggagccctctgctgcacgacagggggccacatgccctacacagatgctgtg	1058							
Db	504	GGCCACAGCTCCCAAGCTCTGTACAGGAACCGAATGCAGTGCCTCTGCTCATGGCCACC	445							
Qy	1059	gtcacgaaggtccagagatacatcgacctgaatccccaccagcctgccccatcgagtgacc	1118							
Db	444	ATTGCCGAGGTGCTGCGTTTGGCGGCTGTGGTGGCCCTTGGCCCTTGGCCCATCTGTGCAACT	385							
Qy	1119	tgtacgtttaactcagaactactctcatcccaagggcacacaactattaaactccctc	1179							
Db	384	AGGCTACGACGATCTCCGGCTATGACATCCCTAAGGATATGTGTATCATCCCCAACATC	325							
Qy	1179	actctgtgtacatgacaacaagaatttccccaccagagatgtttgacctctgtcac	1238							
Db	324	CAAGCGCCAACTTGGATGAGATGGTTTGGGAAGTGCCTCAGCAAGATTCTGCCAGATCGC	265							
Qy	1239	ttcttgatgaagtggaaatttlaagaaaagtaactacttcatgcctttctcagaagga	1298							
Db	264	TTCTGTGAACCTGGGAAGAAATCCCAAGAACACCATCTTTGGCTG-----TGGG	217							
Qy	1299	aaacggatttgtgggagaggccctggcccgcatggagctgttttatcttgaccttc	1358							
Db	215	GCACGCTGTGCTTGGGAGAGCCCTCTGGCAGCGCTGGAGCTCTTTTGGTCTGGCTCGT	157							
Qy	1359	atttacagaactttaacctgaaatctc	1386							
Db	156	CTGTGTCAGGCTTTCACCTCTGCTGCTC	129							

```

RESULT 8
US-106-698-368/c
: Sequence 358, Application US/10106698
: GENERAL INFORMATION:
: APPLICANT: Ruben et al.
: TITLE OF INVENTION: Colon and Colon Cancer
: FILE REFERENCE: PA005P1
: CURRENT APPLICATION NUMBER: US/10/106,698
: CURRENT FILING DATE: 2002-03-27
: PRIOR APPLICATION NUMBER: PCT/US00/26524
: PRIOR FILING DATE: 2000-09-28
: PRIOR APPLICATION NUMBER: US 60/157,137
: PRIOR FILING DATE: 1999-09-29
: PRIOR APPLICATION NUMBER: US 60/163,280
: PRIOR FILING DATE: 1999-11-03
: NUMBER OF SEQ ID NOS: 8564
: SOFTWARE: PatentIn Ver. 3.0
: SEQ ID NO 368

```

```
; LENGTH: 2026
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1326)..(1326)
; OTHER INFORMATION: n equals a,t,g, or c
US-10-106-698-368

Query Match      2.8%; Score 49; DB 6; Length 2026;
Best Local Similarity 43.0%; Pred. No. 0.002;
Matches 238; Conservative 0; Mismatches 315; Indels 0; Gaps 0;

Qy 849 ttcactatgaaacctgttaactcactgagctgaacttaacttaactgagctggaacagagaa 908
    || || || || || || || || || || || || || || || || || || || || || ||
Db 862 ttgactgtggatgagattgtggccaggcccttcaattcttcttctggtggtatgaaatc 803

Qy 909 acaagcacaacccctgagatatactctctctctctctctctctctctctctctctctct 968
    || || || || || || || || || || || || || || || || || || || || || ||
Db 802 ATCACCAACACACTTCTTTTGGCCACTTACTTACTTGGCCACCAACCTGACGTGCAAGAG 743

Qy 969 aaagtcaggaagatgaaacgtgctcattggaagaaacccgagccctgcatgcagac 1028
    || || || || || || || || || || || || || || || || || || || || || ||
Db 742 AAGCTTCTGAGAGAGGTAGACGTTTTTAAGGAGAAACACATGGCCCTGAGTTCTGCAGC 683

Qy 1029 aggggccacatccctacacagatgctgtgtgacagaggtccagagatacatgacctc 1088
    || || || || || || || || || || || || || || || || || || || || || ||
Db 684 CTCGAGGAAGGCGTGCCTTATCTGGACATGGTGATTCAGAGACCCCTGAGGATGTACCG 623

Qy 1089 atccccaccagctgccccatgcagtcagcttgagcttgaattcagaactacctcatt 1148
    || || || || || || || || || || || || || || || || || || || || || ||
Db 622 CCAGCTTTCAGATTACACAGGGAGGAGCAGTCAGGACTGCGAGGTGCGGGCAGCGCATC 563

Qy 1149 ccaaggggcaacacattatacttccctcacttctgtctactatgacaaacagaattt 1208
    || || || || || || || || || || || || || || || || || || || || || ||
Db 562 CCCGAGGCGCTGTGCTAGAGATGCGCGTGGGTGCGCTTGCACCATGACCCCTGAGCATT 503

Qy 1209 ccaacccagagatgtttgaacctcgtcacttctgtgagatgaaggtggaattttaagaa 1268
    || || || || || || || || || || || || || || || || || || || || || ||
Db 502 CCAGGCCCGGAGACCTTCACCCCTGAAAGTTTACGGCTGAGGCCCGGCAGCAGCACCGG 443

Qy 1269 agtaactacttactgcttctcagcaggaacacggaatttgtgtggagagggcctgccc 1328
    || || || || || || || || || || || || || || || || || || || || || ||
Db 442 CCCTTCACGTACTGCGCTTTCGGGGCGGCGCCAGCAGAGCTGCTCGGGGTGCGTCTAGGG 383

Qy 1329 cgcagtagctgttttttctctctcacttctatttaccagaacttttaacctgaattcttg 1388
    || || || || || || || || || || || || || || || || || || || || || ||
Db 382 CTCGTTGAGGTCAGTTGACACTGCTCCAGGTGCTGCACAAAGTTCGGGTTCACAGCCCTGC 323

Qy 1389 attgaccccaagg 1401
    || || || || ||
Db 322 CCGAGACCCAGG 310

RESULT 9
US-10-105-299-7201/c
; Sequence 7201, Application US/10105299
; GENERAL INFORMATION:
; APPLICANT: Rosen, et. al
; TITLE OF INVENTION: Human Secreted Proteins
; FILE REFERENCE: PS950
; CURRENT APPLICATION NUMBER: US/10/105,299
; CURRENT FILING DATE: 2002-03-26
; NUMBER OF SEQ ID NOS: 15197
; Prior Application removed - See File Wrapper or Palm
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7201
; LENGTH: 13327
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-105-299-7201

Query Match      2.6%; Score 45.8; DB 1; Length 1530;
Best Local Similarity 44.6%; Pred. No. 0.012;
Matches 225; Conservative 0; Mismatches 277; Indels 3; Gaps 1;

Qy 882 gacttaacttgagctggagacagacacaaagcacaacccctgagatatactctctctc 941
    || || || || || || || || || || || || || || || || || || || || || ||
Db 901 gatactgtctgtgtgaacacagagattcagcagtcagagtgagtgaggagtggaatttcgag 960

Qy 942 ctgctgaagcaccagaggtcacagctaaagtccagaaagagattgaacgtgtcatggc 1001
    || || || || || || || || || || || || || || || || || || || || || ||
Db 961 ctgctaaagagccagagattttcaaaaagcctacagaagaattggatcgagtaattggg 1020

Qy 1002 agaaacccggagccctgcatgcacagcagggggccacatgcccctcacagatgctgtggtg 1061
    || || || || || || || || || || || || || || || || || || || || || ||
Db 1021 cagaatagatgggtacaagaaaagacattccaaatctctctacatagaggcaatagtc 1080

Qy 1062 cagaggtccagagatacatcgacctcatcccccacccctgcccctgagtgacctgt 1121
    || || || || || || || || || || || || || || || || || || || || || ||
Db 1081 aaagagactatgcactgaccccggtggcaccatattgtgtgcccagggagtgctcgagaa 1140

Qy 1122 gactttaaatcagaataactcctcatctcccaaggggcacaaccattataacttccctcact 1181
    || || || || || || || || || || || || || || || || || || || || || ||
Db 1141 gactgtaaggtagcagcgctacgacgtttaagaaaggaacacagggtccctgtgtgagcgtatg 1200

Qy 1182 tctgtgctacatgacaacaaagaatttcccaaccagagatgtttgacctcgtcacttt 1241
    || || || || || || || || || || || || || || || || || || || || || ||
Db 1201 actatggagagacccctcacatttgggacgagcgtcgagcgttcaagccgagagaggttc 1260

Qy 1242 ctggatgaaggtggaattttaagaaaagtaact---acttcagctcttctcagcagga 1298
    || || || || || || || || || || || || || || || || || || || || || ||
Db 1261 cacaagaaagtcactgatgtttaaaggacatgattttgagcttttgcctatttggagctggg 1320

Qy 1299 aaacggalltgtgtggagagggcctggtcccgcaatgagctgttttttctcctgaccttc 1358
```

Db 1321 agaaggatgcccgggttaataacttggggcttaagggtgattcaagctagcttagctaat 1380
Qy 1359 atttaccagaactttaaccgtgaat 1383
Db 1381 ctatatacattgaatttaactggtcat 1405

RESULT 11

PCT-US02-06912-8

: Sequence 8, Application PC/TUS0206912

: GENERAL INFORMATION:

: APPLICANT: University of Kentucky Research Foundation

: TITLE OF INVENTION: Cytochrome P450s and Uses Thereof

: FILE REFERENCE: 07678/100W03

: CURRENT APPLICATION NUMBER: PCT/US02/06912

: PRIOR FILING DATE: 2002-03-08

: PRIOR FILING DATE: 2001-03-09

: PRIOR FILING DATE: 2001-03-13

: NUMBER OF SEQ ID NOS: 43

: SOFTWARE: FastSeq for Windows Version 4.0

: SEQ ID NO 8

: LENGTH: 1693

: TYPE: DNA

: ORGANISM: Nicotiana tabacum

: PCT-US02-06912-8

Query Match 2.6%; Score 45.8; DB 1; Length 1693;
Best Local Similarity 51.7%; Pred. No. 0.012;
Matches 104; Conservative 0; Mismatches 97; Indels 0; Gaps 0;

Qy 87 aaactccctctggccacactctctccagtgattggaaatactctacagatagatt 145
Db 194 aactccacactgggtcccaacagcagtcctcaattttgtgaattgcttcaagtggc 253
Qy 147 aaggatgtcagcaaatcttaacaaatctctcaaaatctatggccctgtgttcaactcg 206
Db 254 gactgaaccataactctctggccaccatgtcacaaactatgtctctattttactc 313
Qy 207 tatttggcctggaacgcatggtgtgtctgcatgatatgaagtgggtgaaggaagccctg 265
Db 314 aaacttggttcaaaaactagctgtgtatcgaaacctgagctagctaaacaaagtctta 373
Qy 267 attgatctgtgagaggagttt 287
Db 374 cacacgcaagggggtcgagttt 394

RESULT 12

US-09-975-254-18019

: Sequence 18019, Application US/09975254

: GENERAL INFORMATION:

: APPLICANT: Byrum, Joseph R.

: APPLICANT: Heck, Gregory R.

: APPLICANT: La Rosa, Thomas J.

: TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with

: FILE REFERENCE: 38-21(15309)B

: CURRENT APPLICATION NUMBER: US/09/975,254

: PRIOR FILING DATE: 2001-10-12

: PRIOR FILING DATE: 1999-03-05

: NUMBER OF SEQ ID NOS: 31255

: SEQ ID NO 18019

: LENGTH: 254

: TYPE: DNA

: ORGANISM: Glycine max

: OTHER INFORMATION: Clone ID: 700955037H1

: US-09-975-254-18019

Query Match 2.6%; Score 44.6; DB 5; Length 254;
Best Local Similarity 50.2%; Pred. No. 0.011;
Matches 110; Conservative 0; Mismatches 109; Indels 0; Gaps 0;

Qy 1035 acatgccctacacagatgctgggtgacgaggtccagagatacatcgacctcatcccca 1095
Db 6 aaatgccatacctacagctgtgtggaaggagcttctaagaagcaccctccaacact 65
Qy 1096 ccagcctgccccatgcagtgacctgtgacgtttaaatccagaacactcatctcccaagg 1155
Db 66 ttgtgctaacaacatgctgtgactgagccaccacttgggagggtatgacataccaattg 125
Qy 1156 gcacaaccattataacttccctcacttctgtctacatgacacaaagaatttcccaacc 1215
Db 126 atgcaaatgttaggtgtacacacagccatgtcgtgaggagcccaaaaaattgtttaaac 185
Qy 1216 caagatgtttgaacctctcacttctctgtgagtgaaaggtg 1254
Db 186 ctgagaagtttgacctgagagattcatctctgggggtg 224

RESULT 13

US-10-103-313-626

: Sequence 626, Application US/10103313

: GENERAL INFORMATION:

: APPLICANT: Rosen et al.

: TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

: FILE REFERENCE: P207C1

: CURRENT APPLICATION NUMBER: US/10/103,313

: CURRENT FILING DATE: 2002-03-12

: NUMBER OF SEQ ID NOS: 653

: SOFTWARE: PatentIn Ver. 2.0

: SEQ ID NO 626

: LENGTH: 832

: TYPE: DNA

: ORGANISM: Homo sapiens

: US-10-103-313-626

Query Match 2.5%; Score 43.6; DB 6; Length 832;
Best Local Similarity 58.6%; Pred. No. 0.033;
Matches 95; Conservative 0; Mismatches 64; Indels 3; Gaps 1;

Qy 823 agaaagaaacaaacacagctctgaattcactattgaaaacttggtaactcactgcagctg 882
Db 627 agtaacagcagggaccggagagccatttccaggagtagacgtcgtaataacacacgatt 686
Qy 883 acttacttgagctggagacagacaaacacacccctgagatagctctctctctcc 942
Db 687 tttttttgg---cgtcaccgaaacacagcagcaccctgtgctatgggctctctcattc 743
Qy 943 tgcagaagcaccagaggtcacagctaaagtccagggaagaga 984
Db 744 tgcctaagctacctaagaggtgagagtggtgcaacccggagaga 785

RESULT 14

US-10-103-313-625

: Sequence 625, Application US/10103313

: GENERAL INFORMATION:

: APPLICANT: Rosen et al.

: TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

: FILE REFERENCE: P207C1

: CURRENT APPLICATION NUMBER: US/10/103,313

: CURRENT FILING DATE: 2002-03-12

: NUMBER OF SEQ ID NOS: 653

: SOFTWARE: PatentIn Ver. 2.0

: SEQ ID NO 625

: LENGTH: 1001

: TYPE: DNA

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 19, 2002, 08:15:21 ; Search time 277.57 seconds
(without alignment)
2294.890 Million cell updates/sec

Title: US-09-763-292-2
Perfect score: 743
Sequence: 1 tcgaataatttgagcctg.....ttagctcatgtgaacgggg 743

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_1101.*
1: /SIDS2/gcgdata/geneseq/geneseq/NA1980.DAI.*
2: /SIDS2/gcgdata/geneseq/geneseq/NA1981.DAI.*
3: /SIDS2/gcgdata/geneseq/geneseq/NA1982.DAI.*
4: /SIDS2/gcgdata/geneseq/geneseq/NA1983.DAI.*
5: /SIDS2/gcgdata/geneseq/geneseq/NA1984.DAI.*
6: /SIDS2/gcgdata/geneseq/geneseq/NA1985.DAI.*
7: /SIDS2/gcgdata/geneseq/geneseq/NA1986.DAI.*
8: /SIDS2/gcgdata/geneseq/geneseq/NA1987.DAI.*
9: /SIDS2/gcgdata/geneseq/geneseq/NA1988.DAI.*
10: /SIDS2/gcgdata/geneseq/geneseq/NA1989.DAI.*
11: /SIDS2/gcgdata/geneseq/geneseq/NA1990.DAI.*
12: /SIDS2/gcgdata/geneseq/geneseq/NA1991.DAI.*
13: /SIDS2/gcgdata/geneseq/geneseq/NA1992.DAI.*
14: /SIDS2/gcgdata/geneseq/geneseq/NA1993.DAI.*
15: /SIDS2/gcgdata/geneseq/geneseq/NA1994.DAI.*
16: /SIDS2/gcgdata/geneseq/geneseq/NA1995.DAI.*
17: /SIDS2/gcgdata/geneseq/geneseq/NA1996.DAI.*
18: /SIDS2/gcgdata/geneseq/geneseq/NA1997.DAI.*
19: /SIDS2/gcgdata/geneseq/geneseq/NA1998.DAI.*
20: /SIDS2/gcgdata/geneseq/geneseq/NA1999.DAI.*
21: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAI.*
22: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAI.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed. and is derived by analysis of the total score distribution...

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	622.6	83.8	8437	AAD12242	Human cytochrome P
2	433.6	58.4	1001	AAH51158	Human CYP2C8 relat
3	421.8	56.8	1001	AAH51159	Human CYP2C8 relat
4	415.4	55.9	436	AAH51164	Human CYP2C9 relat
5	369	49.7	734	AAD12247	Human cytochrome P
6	366.2	49.3	557	AAD12245	Human cytochrome P
7	193.2	26.0	570	AAD12249	Human cytochrome P
8	163	21.9	1419	AAH51158	Human cytochrome P
9	163	21.9	1419	AAH51158	Human cytochrome P
10	163	21.9	1473	AAH51158	Human cytochrome P
11	163	21.9	1473	AAH51158	Human derived cyto

12 163 21.9 1591 20 AAX22715 Human cytochrome P
13 163 21.9 1845 20 AAX19925 Mammalian cytochro
14 163 21.9 1852 19 AAT11381 Cytochrome P450 2C
15 163 21.9 1852 19 AAV44157 Human cytochrome P
16 163 21.9 1854 17 AAT11378 Human cytochrome P
17 163 21.9 1854 19 AAV44154 Human cytochrome P
18 163 21.9 2829 21 AAQ93333 Partial sequence o
19 161.4 21.7 1818 11 AAQ03599 Human liver cytoch
20 153.4 20.6 1892 19 AAV44159 Human cytochrome P
21 150.6 20.3 169 22 AAD12246 Human cytochrome P
22 149.6 20.1 1419 22 AAF76959 Human derived cyto
23 149.6 20.1 1438 22 AAF76956 Human derived cyto
24 149.6 20.1 1447 19 AAV12758 Cytochrome P450 2C
25 149.6 20.1 1473 16 AAQ87728 Human auxillary cy
26 149.6 20.1 1473 17 AAT28394 Human cytochrome P
27 149.6 20.1 1669 19 AAV12759 Wild type cytochro
28 149.6 20.1 1746 17 AAT03708 Cytochrome P450 2C
29 149.6 20.1 1746 19 AAV44153 Human cytochrome P
30 149.6 20.1 1748 22 AAD12241 Human cytochrome P
31 149.6 20.1 8475 22 AAF76964 Genetic construct
32 144 19.4 200 16 AAQ94327 Human cytochrome P
33 140.2 18.9 1473 16 AAQ87727 Human auxillary cy
34 140.2 18.9 1473 17 AAT28393 Human cytochrome P
35 140.2 18.9 2009 17 AAT11379 Cytochrome P450 2C
36 140.2 18.9 2009 19 AAV44155 Human cytochrome P
37 140.2 18.9 2258 17 AAT11380 Cytochrome P450 2C
38 140.2 18.9 2258 19 AAV44158 Human cytochrome P
39 140 18.8 167 22 AAD12248 Human cytochrome P
40 135.8 18.3 1473 16 AAQ87724 Human auxillary cy
41 135.8 18.3 1473 17 AAT28390 Human cytochrome P
42 135.8 18.3 1473 17 AAT17394 Human derived cyto
43 135.8 18.3 1922 22 AAH57473 Human liver cell s
44 134.2 18.1 1473 16 AAQ87725 Human auxillary cy
45 134.2 18.1 1473 16 AAQ87726 Human auxillary cy

ALIGNMENTS

RESULT 1
AAD12242
ID AAD12242 standard; DNA: 8437 BP.
XX
AC AAD12242:
XX
XX 25-SEP-2001 (first entry)
XX Human cytochrome P450 (CYP450) 2C19 gene.
XX Human: gene structure: phenotypic expression; guanosine cofactor;
KW germline variation analysis; exon-intron boundary; Tetrahymena rRNA;
KW cytochrome P450 2C19; CYP450 2C19; ds.
XX Homo sapiens.
OS
XX
XX Key Location/Qualifiers
FH misc_signal 700..702
FH /tag= a
FH primer_bind /note= "Translation start site"
FH 818..833
FH /tag= b
FH /bound_moiety= "Primer"
FH 1535..1557
FH /tag= c
FH /bound_moiety= "Primer"
FH primer_bind 1578..1597
FH /tag= d
FH /bound_moiety= "Primer"
FH primer_bind 1861..1882
FH /tag= e
FH /bound_moiety= "Primer"
FH primer_bind 1879..1898
FH /tag= f

FT	primer_bind	/bound_moiety=	"Primer"
FT		3147..3168	
FT		/*tag= g	
FT	primer_bind	/bound_moiety=	"Primer"
FT		3258..3279	
FT		/*tag= h	
FT	primer_bind	/bound_moiety=	"Primer"
FT		3460..3479	
FT		/*tag= i	
FT	primer_bind	/bound_moiety=	"Primer"
FT		4876..4890	
FT		/*tag= j	
FT	primer_bind	/bound_moiety=	"Primer"
FT		6049..6064	
FT		/*tag= k	
FT	primer_bind	/bound_moiety=	"Primer"
FT		6091..6105	
FT		/*tag= l	
FT	primer_bind	/bound_moiety=	"Primer"
FT		6093..6112	
FT		/*tag= m	
FT	primer_bind	/bound_moiety=	"Primer"
FT		6858..6872	
FT		/*tag= n	
FT	primer_bind	/bound_moiety=	"Primer"
FT		6874..6893	
FT		/*tag= o	
FT	primer_bind	/bound_moiety=	"Primer"
FT		7808..7827	
FT		/*tag= p	
FT	primer_bind	/bound_moiety=	"Primer"
FT		7810..7831	
FT		/*tag= q	
FT	primer_bind	/bound_moiety=	"Primer"
FT		7937..7956	
FT		/*tag= r	
FT	primer_bind	/bound_moiety=	"Primer"
FT		7940..7961	
FT		/*tag= s	
FT	primer_bind	/bound_moiety=	"Primer"
FT			

WO200153529-A2.

26-JUL-2001.

17-JAN-2001: 2001WO-US01461.

20-JAN-2000- 2000US-0488127

(GENO-) GENOME THERAPEUTICS CORP.

Thomann H. Fitzgerald MS.

WPT: 2001-465380/50

Determining structure of genes whose sequence is not known from cDNA, by sequencing the gene or gene across exon-intron boundaries using evenly spaced primers comprising nucleic acids that hybridize to the cDNA of gene -

Claim 23: Fig 5: 81pp: English.

The present invention relates to a method for determining gene structure when the genomic sequence is unknown. The method involves sequencing the gene across exon-intron boundaries using evenly spaced primers or tiled primers. The tiled primers comprises nucleic acids that hybridise to the known cDNA sequence of the gene at about 100-300 base intervals and the gene comprises the template. Gene structure can be determined without the need to sequence the entire gene. The method provides information necessary to determine gene structure and phenotypic expression without the need to sequence entire chromosomal copy of the gene or fragment. The methods are useful in germline sequence variation analysis. The method is also useful for determining the boundaries

between regions of nucleic acids that were separated by intervening sequence, and also for determining boundaries present in genes containing group 1 type introns such as Tetrahymena rRNA, where self-splicing occurs in the presence of guanosine cofactor. The present sequence is human cytochrome P450 (CYP450) 2C19 gene related to the invention.

Query Match	83.8%	Score 622.6;	DB 22;	Length 8437;
Best Local Similarity	93.3%	Pred. No. 1.3e-183;		
Matches 694;	Conservative	0;	Mismatches 45;	Indels 5;
Gaps 4				

[illegible]

RESULT 2

AAH51158
ID AAH

XXXXXXXXXXXXXXXXXXXX

XX

1007 NOV 27 1966


```
XX PF 24-MAR-2000; 2000WO-IB00403.
XX PR 25-MAR-1999; 99US-0126269.
XX PR 30-APR-1999; 99US-0131961.
XX PA (GEST ) GENSET.
XX PI Blumenfeld M, Bougueleret L, Chumakov I, Cohen-Akenine A:
XX DR WPI: 2000-638353/61.
XX PI Polynucleotides comprising sequences from malate decarboxylase
XX PT enzyme-related biallelic markers used for genotyping -
XX PS Claim 13; Page 280; 673pp; English.
XX CC Sequences AAH51110-AAH51593 represent human DNA fragments which contain
XX CC biallelic markers. The sequences are related to various human genes
XX CC including microsomal glutathione S-transferase II (MGSTII), malate
XX CC decarboxylase enzyme (DME1/ME1), cytochrome P450, glutathione
XX CC reductase/synthase (GSHR/GSHS), flavin-containing monooxygenases (FMO),
XX CC gamma-glutamyltransferase 5 (GGT5), dipeptidase (DP), glucose 5-phosphate
XX CC dehydrogenase (G6PDH), phosphogluconate dehydrogenase (PGDH), and uridine
XX CC diphosphate glucuronosyl transferases (UGT2). Each of these sequences
XX CC contains a biallelic marker/polymorphism, which is represented in the
XX CC sequence as a degenerate/undefined base. The genes to which the biallelic
XX CC marker containing sequences are related are involved in drug metabolism.
XX CC Sequences AAH51594 - AAH51598 represent the genomic sequence of the
XX CC MGSTII gene and four alternative MGSTII cDNA sequences. AAH52905-AAH52906
XX CC are MGSTII gene products. PCR primers AAH51599 and AAH51500 are used in
XX CC an example for the amplification of human genomic DNA fragments. The
XX CC invention includes a method of genotyping comprising determining the
XX CC identity of a nucleotide at a DME- or MGSTII-related biallelic marker in
XX CC a biological sample. The method is used to determine the frequency in
XX CC a population of an allele of a DME- or MGST-II related biallelic marker and
XX CC to select an individual for inclusion in a clinical trial of a drug
XX CC treatment. The method is also used to detect association between allele
XX CC and phenotype, and to detect association between haplotype and phenotype.
XX CC The polynucleotides are used, in hybridization assays, sequencing assays
XX CC or allele specific amplification assays. The method can be used to
XX CC determine whether an individual suffers or is at risk of developing
XX CC asthma or is at risk of developing hepatotoxicity on treatment with
XX CC zileuton.
XX SO Sequence 1001 BP; 265 A; 188 C; 244 G; 302 T; 2 other;

Query Match 56.8%; Score 421.8; DB 21; Length 1001;
Best Local Similarity 78.0%; Pred. No. 2.1e-121;
Matches 544; Conservative 1; Mismatches 148; Indels 4; Gaps 3;

Qy 49 atgaaaataatcagctaaatcaggcttagcaaatgagcaaaaatagtaacttccttgcctg 108
Db 1 atgaaaatgcagctgaatcacaatagctatttgagcaaaatagcgaacttatttgcctg 60

Qy 109 ttatctgtctacttctcctagctctcaaaaggctctatggccctgtgttcaactctgatt 168
Db 61 ctattgtcatttctccctcagttctcaaaagctctatggctcctgtgttcaactctgatt 120

Qy 169 tggcctgaaccataatggtgtgtgatgatgatgaatgaatgaagaagccctattga 228
Db 121 tggcatgaatcccatagtggtgttctatgatgatgatgagcagtgaaagaaagccctattga 180

Qy 229 tcttggaggagagttctctgaaagagcatttcccaactgagctgaaagagcctaagagg 288
Db 181 taatggagagaggtttcttggagagagcaaatccccaataatctcaagaagaattacaaagg 240

Qy 289 atttggtaggtgtgctgctgttctcagcatctgtcttgggagatgggagatggaaa 348
Db 241 acttggtaggtgtgcatatttctgtgtcagcttggtaactggggtgaggggagatggaaa 300

Qy 349 acagagacttacagagctctctcgggagagacttggcccaccatgctgcccctagt 407
```

```
Db 301 acagagcctaaaaagctcttc-agcagagcttagcctatctgcagctgcccagtgct 359
Qy 408 -cagcttctcttcttctgctgggagctcctcctcctagttcttcttctctctcttagga 466
Db 350 gcagcaacttcttcttctgctggaattctc-ccagtttctgccccctttttttattagga 418
Qy 457 attgttttcagaacaatgaaagaaatgaaagagatccggcgtttctcctcctcagcctg 526
Db 419 atcatttccagcaatggaagagatggaagagatccggcgtttctcctcctcacaaccttg 478
Qy 527 cgaatatttggagtggaagagcattgagacccgttgcagacccgttccaagagagagcccgctgc 586
Db 479 cgaatatttggagtggaagagcattgagacccgttgcagacccgttccaagagagagcctcgc 538
Qy 587 ctgttgaggaggttgagaaaaaaccaagaggtgggtgagccctactcactacactacactta 646
Db 539 ctgttgaggaggttgagaaaaaaccaagaggtgggtgagctctactctgcctcattgacctta 598
Qy 647 ctggactactatcttctactgacattcttggaacatttcagggggtggccatcttt 706
Db 599 acagttactcttctcactagtagcagctccttggaacaatttcagggggtggccaggtcttc 658
Qy 707 cattatgagctctggttcttagctcatgtgaaagcgggg 743
Db 659 attgcgcactcctggttctcagccctcaggtcaggtgag 695

RESULT 4
AAH51164
ID AAH51164 standard; DNA; 436 BP.
XX AC AAH51164;
XX DT 29-AUG-2001 (first entry)
XX DE Human CYP2C9 related DNA containing a biallelic polymorphism SEQ ID 55.
XX KW Human; biallelic marker; single nucleotide polymorphism; SNP; MGSTII;
XX KW microsomal glutathione S-transferase II; malate decarboxylase enzyme;
XX KW DME1; ME1; cytochrome P450; glutathione reductase; GSHR; GSHS; GGT5;
XX KW flavin-containing monooxygenase; FMO; gamma-glutamyltransferase 5;
XX KW dipeptidase; DP; glucose 6-phosphate dehydrogenase; G6PDH; haplotype;
XX KW phosphogluconate dehydrogenase; PGDH; drug metabolism; phenotype;
XX KW uridine diphosphate glucuronosyl transferase; UGT2; asthima; hepatotoxicity;
XX KW zileuton; ds.
XX OS Homo sapiens.
XX PN WO200058508-A2.
XX PD 05-OCT-2000.
XX PF 24-MAR-2000; 2000WO-IB00403.
XX PR 25-MAR-1999; 99US-0126269.
XX PR 30-APR-1999; 99US-0131961.
XX PA (GEST ) GENSET.
XX PI Blumenfeld M, Bougueleret L, Chumakov I, Cohen-Akenine A:
XX DR WPI: 2000-638353/61.
XX PI Polynucleotides comprising sequences from malate decarboxylase
XX PT enzyme-related biallelic markers used for genotyping -
XX PS Claim 13; Page 284; 673pp; English.
XX CC Sequences AAH51110-AAH51593 represent human DNA fragments which contain
XX CC biallelic markers. The sequences are related to various human genes
XX CC including microsomal glutathione S-transferase II (MGSTII), malate
XX CC decarboxylase enzyme (DME1/ME1), cytochrome P450, glutathione
```

CC reductase/synthase (GSHR/GSHS), flavin-containing monooxygenase (FMO),
CC gamma-glutamyltransferase 5 (GGT5), dipeptidase (DP), glucose 6-phosphate
CC dehydrogenase (G6PDH), phosphogluconate dehydrogenase (PGDH), and uridine
CC diphosphate glucuronosyl transferases (UGT). Each of these sequences
CC contains a biallelic marker/polymorphism, which is represented in the
CC sequence as a degenerate/undefined base. The genes to which the biallelic
CC marker containing sequences are related are involved in drug metabolism.
CC Sequences AAH51594 - AAH51598 represent the genomic sequence of the
CC MGSTII gene and four alternative MGSTII cDNA sequences. AAB52905-
CC are MGSTII gene products. PCR primers AAH51599 and AAH51600 are used in
CC an example for the amplification of human genomic DNA fragments. The
CC invention includes a method of genotyping comprising determining the
CC identity of a nucleotide at a DME- or MGSTII-related biallelic marker in
CC a biological sample. The method is used to determine the frequency in
CC population of an allele of a DME- or MGSTII-related biallelic marker and
CC to select an individual for inclusion in a clinical trial of a drug
CC treatment. The method is also used to detect association between allele
CC and phenotype, and to detect association between haplotype and phenotype.
CC The polynucleotides are used, in hybridization assays, sequencing assays,
CC or allele specific amplification assays. The method can be used to
CC determine whether an individual suffers or is at risk of developing
CC asthma or is at risk of developing hepatotoxicity on treatment with
CC zileuton.
XX
SQ Sequence 436 BP; 100 A; 94 C; 105 G; 124 T; 13 other;

Query Match 55.9%; Score 415.4; DB 21; Length 456;
Best Local Similarity 96.3%; Pred. No. 1.3e-119;
Matches 421; Conservative 13; Mismatches 2; Indels 1; Gaps 1;
QY 22 gtgctgaataaagacatacaatacaatgaaataatcatgctaaatcaggcttgacaaa 81
DB 1 gtgctgaataaagacatacaatacaatgaaataatcatgctaaatcaggcttgacaaa 50
QY 82 tggacaaaatagtaactctgttctgtctctctctctctctctctctctctctctctcaagct 141
DB 61 tggacaaaatagtaactctgttctgtctctctctctctctctctctctctctctctcaagct 120
QY 142 ctatgcccctgtctcactctgtatttttgccctgaaacccatagtgctgctcaatgata 201
DB 121 ctatgcccctgtctcactctgtatttttgccctgaaacccatagtgctgctcaatgata 180
QY 202 tgaacagtgaaagagccctgattgattcttgagagagagattcttgagagagagatttt 251
DB 181 tgaaghrtgaaagagccctgattgattcttgagagagagattttctgagagagagatttt 240
QY 262 cccactggtgaaagagctaacagagatttgtaggtgtgcatgtcctctgttttagcat 321
DB 241 cccactggtgaaagagctaacagagatttgtaggtgtgcatgtcctctgttttagcat 300
QY 322 ctgtcttgagagtgaggagtgagagtgagagtgagagtgagagtgagagtgagagtgag 381
DB 301 ctgtcttgagagtgaggagtgagagtgagagtgagagtgagagtgagagtgagagtgag 350
QY 382 ggcctccacatggctgcccagtgccagcttctctctctctctctctctctctctctctct 441
DB 361 ggcctccacatggctgcccagtgccagcttctctctctctctctctctctctctctct 419
QY 442 agttctgttct 458
DB 420 agttctgttct 436
RESULT 5
AAD12247/c
ID AAD12247 standard; DNA: 734 BP.
XX
AC AAD12247;
XX
DT 25-SEP-2001 (first entry)
XX
DE Human cytochrome P450 (CYP450) 2C19 gene PCR product: #3.

XX Human: gene structure; phenotypic expression; guanosine cofactor;
KW germline variation analysis; exon-intron boundary; Tetrahymena rRNA;
KW cytochrome P450 2C19; CYP450 2C19; ds.
XX
XX Homo sapiens.
OS
PN W0200153529-A2.
XX
PD 26-JUL-2001.
XX
PF 17-JAN-2001; 2001WO-US01461.
XX
PR 20-JAN-2000; 2000US-0488127.
XX
PA (GENO-) GENOME THERAPEUTICS CORP.
XX
PI Thomann H, Fitzgerald MS;
XX
DR WPI: 2001-465380/50.
XX
PI Determining structure of genes whose sequence is not known from cDNA,
PI by sequencing the gene or gene across exon-intron boundaries using
PI evenly spaced primers comprising nucleic acids that hybridize to the
PI cDNA of gene
XX
PS Example 2; Fig 4; 8lpp; English.
XX
CC The present invention relates to a method for determining gene
CC structure when the genomic sequence is unknown. The method involves
CC sequencing the gene across exon-intron boundaries using evenly spaced
CC primers or tiled primers. The tiled primers comprises nucleic acids that
CC hybridize to the known cDNA sequence of the gene at about 100-300 base
CC intervals and the gene comprises the template. Gene structure can be
CC determined without the need to sequence the entire gene. The method
CC provides information necessary to determine gene structure and phenotypic
CC expression without the need to sequence entire chromosomal copy of the
CC gene or fragment. The methods are useful in germline sequence variation
CC analysis. The method is also useful for determining the boundaries
CC between regions of nucleic acids that were separated by intervening
CC sequence, and also for determining boundaries present in genes containing
CC group 1 type introns such as Tetrahymena rRNA, where self-splicing occurs
CC in the presence of guanosine cofactor. The present sequence is a PCR
CC product obtained from human cytochrome P450 (CYP450) 2C19 gene
CC and 349L primer related to the invention.
XX
SQ Sequence 734 BP; 236 A; 170 C; 126 G; 199 T; 3 other;
Query Match 49.7%; Score 369; DB 22; Length 734;
Best Local Similarity 90.7%; Pred. No. 5.3e-105;
Matches 427; Conservative 0; Mismatches 40; Indels 4; Gaps 3;
QY 1 tcagaaaattttgaagcctgtgtgctgaaataaagacatacaatacaataatca 50
DB 458 TCAGAAAATATTTGAGCCCTGTGTGACTTTGAATAAAGCATACAAATACAAATGAAATG 409
QY 61 tctaaatcagccttagcaaatggacaaatagtaactctgttctgtctctctctctctct 120
DB 408 ATCTAAGTCAGGCTTAGTAAATGG-CANAAACAGTCACTTCATTTCGTCTAATCTA 350
QY 121 acttctctagctctcaaaaggtctatggccctgtgttctactctgtatttttggcctgaaacc 180
DB 349 CCTTTTCTAGCTCTCAAAAATCTATGGCCCTGTGTTCACCTCTGTATTTTGGCCTGGA 290
QY 181 catagtgtgtctgcacatgaatgaagaacgaagcctgattgattgttgaggaga 240
DB 289 CATGTGGTGTGCTGCAATGGAATATCAAGTGTGAAGGANGCCCTGATTGATCTTGGAGAG 230
QY 241 gtttcttggaagagcattttccctactggctgaaagagctaacagagagatttggtaggtg 300
DB 229 GTTTTCTGGAAGANGCCATTTCCTACTGGCTGCTGAAGAGCTAACAGAGGATTGGTAGGTG 170

XX WPI: 1996-136337/14.
 DR P-PSDB: AAR81465.
 XX
 XX Antibody recognising human derived cytochrome P450C9 - allows
 PT specific detection of cytochrome P450 species in humans
 XX
 XX Example 1: Pages 11-13; 13pp; Japanese.
 XX
 XX The present sequence encodes the human derived cytochrome (HDC)
 CC P450C9, which was obtd. from a commercial cDNA library. Yeast
 CC were transfected with an expression vector contg. the HDC cDNA,
 CC cultured and then disrupted to give a microsomal fraction. The
 CC HDC was purified from the fraction, and used to immunise and
 CC sensitise a mammal. Blood was drawn from the mammal, and an
 CC anti-HDC antibody isolated. The antibody obtd. recognises HDC
 CC P450C9, partic. at a serum dilution rate of 1:10000, and is
 CC substantially without cross reaction to other HDC P450 spp..
 XX
 SQ Sequence 1473 BP: 413 A: 345 C: 319 G: 396 T: 0 other:
 Query Match 21.9%; Score 163; DB 17; Length 1473;
 Best Local Similarity 97.1%; Pred. No. 1.5e-40;
 Matches 166; Conservative 0; Mismatches 5; Indels 0; Gaps 0:
 QY 126 cctagctctcaagggtctatggccctgtgttcactctgtatttggccctgaaccacatag 185
 || |||||
 Db 164 ccaatctctcaagggtctatggccctgtgttcactctgtatttggccctgaaccacatag 223
 QY 186 tggctgtcatgatgatgaagcagtgtaaggaagccctgattgatcttggaggagagtttt 245
 |||||
 Db 224 tggctgtcatgatgatgaagcagtgtaaggaagccctgattgatcttggaggagagtttt 283
 QY 246 ctggaagaggcatttccctcactggctgaagagcctaacagagatttgata 296
 |||||
 Db 284 ctggaagaggcatttccctcactggctgaagagcctaacagagatttgata 334
 RESULT 12
 AAX22715
 ID AAX22715 standard; DNA: 1591 BP.
 XX
 AC AAX22715;
 XX
 DT 27 MAY-1999 (first entry)
 XX
 DE Human cytochrome P450 2C10 DNA.
 XX
 KW Cytochrome P450: detergent; enzyme inhibitor: non-ionic; ionic;
 KW purification; drug oxidation; steroid; carcinogen; pesticide; human; ss.
 XX
 OS Homo sapiens.
 XX
 PN US5886157-A.
 XX
 PD 23-MAR-1999.
 XX
 PF 10-FEB-1994; 94US-0194981.
 XX
 PR 10-FEB-1994; 94US-0194981.
 XX
 PA (UYVA-) UNIV VANDERBILT.
 XX
 PI Gillam EMJ, Guengerich FP, Guo Z, Sandhu P;
 XX WPI: 1999-228609/19.
 DR
 DR Purifying recombinant cytochrome P450 - by utilising novel
 PT combinations of detergents and enzyme inhibitors
 PT
 XX Disclosure: Column 53-54; 91pp; English.
 PS
 XX
 CC This invention describes a recombinant cytochrome P450 protein which
 CC is purified from a host cell culture using a combination of detergents
 CC and enzyme inhibitors. The method comprises (a) fractionating the host
 CC cells to prepare their membranes, (b) adding a non-ionic detergent to
 CC the membranes in a concentration of 0.8% to 2% (w/v) in a
 CC detergent:protein ratio of between 4:1 to 10:1, (c) adding an ionic
 CC detergent to the membranes in a concentration of 0.4% to 0.8% (w/v) in a
 CC detergent:protein ratio of between 2:1 to 4:1 (d) centrifuging the
 CC membrane-detergent mixture to remove insoluble materials and (e)
 CC purifying the protein through a diethylaminoethyl-beaded column, then
 CC through a carboxymethyl-beaded column, and finally through a
 CC hydroxylapatite column. The method is used to purify cytochrome P450
 CC proteins which are responsible for catalysing the oxidation of drugs,
 CC steroids, carcinogens and pesticides. The method simplifies the
 CC purification of P450 proteins, by the use of improved expression vectors
 CC and novel detergent combinations.
 XX
 SQ Sequence 1591 BP: 443 A: 372 C: 353 G: 423 T: 0 other:
 Query Match 21.9%; Score 163; DB 20; Length 1591;
 Best Local Similarity 97.1%; Pred. No. 1.6e-40;
 Matches 165; Conservative 0; Mismatches 5; Indels 0; Gaps 0:
 QY 126 cctagctctcaagggtctatggccctgtgttcactctgtatttggccctgaaccacatag 185
 || |||||
 Db 164 ccaatctctcaagggtctatggccctgtgttcactctgtatttggccctgaaccacatag 223
 QY 186 tggctgtcatgatgatgaagcagtgtaaggaagccctgattgatcttggaggagagtttt 245
 |||||
 Db 224 tggctgtcatgatgatgaagcagtgtaaggaagccctgattgatcttggaggagagtttt 283
 QY 246 ctggaagaggcatttccctcactggctgaagagcctaacagagatttgata 296
 |||||
 Db 284 ctggaagaggcatttccctcactggctgaagagcctaacagagatttgata 334
 RESULT 13
 AAX19925
 ID AAX19925 standard; DNA: 1845 BP.
 XX
 AC AAX19925;
 XX
 DT 11-JUN-1999 (first entry)
 XX
 DE Mammalian cytochrome P450 protein CYP2C9 encoding DNA.
 XX
 KW Bacterial; mammalian: cytochrome P450; chimeric; fusion protein;
 KW oxidise; hydrocarbon; carbon-hydrogen bond; hydroxylating;
 KW bioremediation; environmental pollutant; ss.
 XX
 OS Mammalia.
 XX
 PN WO9908812-A1.
 XX
 PD 25-FEB-1999.
 XX
 PF 17-AUG-1998; 98WO-US16979.
 XX
 PR 20-AUG-1997; 97US-0056754.
 XX
 PA (UYRP) UNIV ROCHESTER.
 XX
 PI Jones JP, Shimoji M;
 XX WPI: 1999-190131/16.
 DR P-PSDB: AAY04127.
 XX
 XX New P450 fusion proteins - comprising a portion of a bacterial
 PT cytochrome P450 protein and a portion of a mammalian cytochrome P450.
 PT protein
 XX
 PS Disclosure: Page 10-11; 51pp; English.

PA (ROMK/) ROMK-SPARKS M.
XX Goldstein JA, Romkes-sparks M;
XX WPI; 1998-436528/37.
XX
XX Screening for drugs metabolised by cytochrome P450 - for identifying
PT mutagenic, carcinogenic, or cytotoxic compounds
XX
XX Example 2; Column 53-56; 63pp; English.
XX
XX This sequence encodes a human cytochrome P450 2C9 polypeptide isolated
CC from clone 65. This polypeptide is used in a method to screen for a drug
CC that is metabolised by a cytochrome P450 having S-mephenytoin
CC 4'-hydroxylase activity. The protein can also be used to identify a
CC mutagenic, carcinogenic or cytotoxic compound. Cytochrome P450 are a
CC large family of haemoprotein enzymes capable of metabolising xenobiotics
CC such as drugs carcinogens and environmental pollutants as well as
CC endobiotics such as steroids, fatty acids and prostaglandins.
XX
SQ Sequence 1852 BP; 514 A; 424 C; 380 G; 534 T; 0 other;

Query Match 21.9%; Score 163; DB 19; Length 1852;
Best Local Similarity 97.1%; Pred. No. 1.7e-40;
Matches 166; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Oy 126 cctagctctcaaggctctatggccctgtgttcactctgtattttggccctgaaccctatag 185
Db 174 ccaatctctcaaggctctatggccctgtgttcactctgtattttggccctgaaccctatag 233

Oy 186 tgggtctgcctgatgaagcagtgaaagagccctgattgattcttgaggagatttt 245
Db 234 tgggtctgcctgatgaagcagtgaaagagccctgattgattcttgaggagatttt 293

Oy 246 ctggaagaggcattttccctctggctgaaagagctaacagagattttgta 296
Db 294 ctggaagaggcattttccctctggctgaaagagctaacagagattttgta 344

Search completed: April 19, 2002, 08:15:28
Job time: 3228 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 19, 2002, 08:07:44 ; Search time 2723.34 Seconds
(without alignments)
2931.736 Million cell updates/sec

Title: US-09-763-292-2

Perfect score: 743

Sequence: 1 tcgaataatttgaagctg,.....ttagctcatgtgaagcggg 743

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 537289281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_hic:*
10: gb_est1:*
11: gb_est2:*
12: gb_hic:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rtd:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	324.2	43.6	675	10	AV683520
2	322.8	43.4	628	10	AV693304
3	322.8	43.4	631	10	AV695514
4	322.8	43.4	692	10	AV697749
5	321.8	43.3	551	10	AV697837
6	321.8	43.3	675	10	AV698435
7	321.8	43.3	677	10	AV697538
8	321.8	43.3	677	10	AV698645
9	317	42.7	580	10	AV696639
10	314.8	42.4	522	10	AV697579
11	289.6	39.0	747	10	AV646984
12	280.2	37.7	645	10	AV696485

13	270.4	36.4	721	11	BG617892
14	267.4	36.0	530	10	AV695838
15	253.2	34.1	395	10	AV660643
16	249.8	33.6	608	11	BG563276
17	245.6	33.1	716	11	BG569317
18	225.8	30.4	425	11	T53876
19	221.4	29.8	724	10	AV697373
20	209.8	28.2	667	10	AV697578
21	171	23.0	401	10	AV694577
22	163	21.9	344	10	AV656121
23	163	21.9	346	10	AV661103
24	163	21.9	360	10	AV659393
25	163	21.9	383	10	AV649200
26	163	21.9	385	10	AV649420
27	163	21.9	388	10	AV655432
28	163	21.9	407	10	AV645669
29	163	21.9	445	10	AV649415
30	163	21.9	612	10	AV661158
31	163	21.9	613	10	AV655841
32	163	21.9	615	10	AV685094
33	163	21.9	620	10	AV647944
34	163	21.9	638	10	AV698137
35	163	21.9	667	10	AV653206
36	162	21.8	361	10	AV661559
37	152	21.8	362	10	AV661295
38	152	21.8	364	10	AV661281
39	155.2	20.9	371	10	AV691273
40	155.2	20.9	376	10	AV691272
41	155.2	20.9	390	10	AV688076
42	155.2	20.9	396	10	AV690361
43	155.2	20.9	409	10	AV687011
44	152.4	20.5	363	10	AV655255
45	152.4	20.5	421	10	AV650504

ALIGNMENTS

RESULT 1

AV683520
LOCUS AV683520 675 bp mRNA EST 25-SEP-2000
DEFINITION AV683520 GK Homo sapiens cDNA clone GKCBJD11 5', mRNA sequence.
ACCESSION AV683520
VERSION AV683520.1 GI:10285383
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 675)
AUTHORS Wu, T., Qian, B., Huang, O., Huang, C., Kang, B., Gao, X., Xu, Z., Xiao, H., Xu, X., Li, N., Peng, Y., Liu, F., Ou, J., Song, H., Cheng, Z., Zeng, L., Yu, S., Gu, W., Tu, Y., Jia, J., Fu, G., Ren, S., Zhong, M., Lu, G., Yang, Y., Gu, Y., Chen, Z. and Han, Z.
Homo sapiens cDNA GK- clones
Unpublished (2000)
Contact: Zeguang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai 201203, P. R. China
Tel: 86-21-50801919(ex.45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.

TITLE

JOURNAL

COMMENT

FEATURES

source

Location/Qualifiers
1..575
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="GKCBJD11"
/clone_lib="GKC"
/tissue_type="hepatocellular carcinoma"
/dev_stage="Adult"
/lab_host="SOLR"

```
/note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
XhoI"
BASE COUNT      178 a   138 c   165 g   194 t
ORIGIN

Query Match      43.6%; Score 324.2; DB 10; Length 675;
Best Local Similarity 78.2%; Pred. NO. 2e-78;
Matches 415; Conservative 0; Mismatches 113; Indels 3; Gaps 2;

Qy 86 caaaatagtaactcgtttgtctgtatctctgtctacttctctagctctcaaaagtctat 145
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5  CACCAATGTAAGTCGCTTAAGTCTCTCCAGCCNAITGCARAGGTCTCAAAAGTCTAT 54

Qy 146 ggcctctgtctactctgtatttttggcctgaaaccccatagtgctgcatgatatgaa 205
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 65 GGTCTGTGTTCACCGTGTATTTGGCATGAATCCCATAGTGTGTTCATGATATGAG 124

Qy 206 qcagtgaaagaaacccctgattgattcttgagagagatttcttgaagagcattttccca 255
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 125 GCAGTGAAGGAAGCCCTGATGATATGAGAGAGAGATTTTCTGGAAGAGCAATCCCA 184

Qy 266 ctgctgaaagagatgaacagagattgttagtgatgctgctcctgtttcagcatctgt 325
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 185 ATATCTCAAGAAATTACTAAGGACTTGGTAGGTGCACATATTTCTGTGACGTTTGG 244

Qy 326 cttgggagtgaggagtgaaacagagacttacagagctcctcctgggcagagccttggcc 385
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 245 AACTGGGGTGAGGGGATGAAACACAGAGCCCTAAAAAGCTTCTC-AGCAGAGCTTAGCC 303

Qy 386 catcacatgctgcccagctgt--cagcttcctctcttctgctggatcctcctcctag 443
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 304 TATCTGATGGCTGCCAGTGTTCAGCAGCTTTCCTCTTGGGTGTGAAATCTCCAGT 363

Qy 444 ttgttctctctctgttaggaattgttttcagcaatggaagaaatggaagagatcc 503
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 364 TCTGCCCTTTTATAGGAATCATTTTCAGCAATGGAAGAGAGATGGAAGGAGATCC 423

Qy 504 ggcattctccctcatagctcagaaatttggatggagagagagacattatgaagacc 563
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 424 GGCCTTCTCCCTCACAACTTCGGAAATTTGGGATGGGAAGAGAGAGCATTCAGGACC 483

Qy 564 gtgtcaagaggaagccgctgctgtgagagagttgagaaacacaaagg 614
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 484 GTGTCAAGGAAGAGCTCACTGCTTGTGGAGGAGTTGAGAAAACCAAGG 534

RESULT 2
AV693304      628 bp   mRNA      EST      25-SEP-2000
LOCUS      AV693304 GK Homo sapiens cDNA clone GKCB04 5', mRNA sequence.
DEFINITION      AV693304
ACCESSION      AV693304
VERSION      AV693304.1 GI:10295167
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 628)
AUTHORS      Wu, T., Qian, B., Huang, O., Huang, C., Kang, B., Gao, X., Xu, Z., Xiao, H.,
, Xu, X., Li, N., Peng, Y., Liu, F., Qu, J., Song, H., Cheng, Z., Zeng, L.,
Xu, S., Gu, W., Tu, Y., Jia, J., Fu, G., Ren, S., Zhong, M., Lu, G., Yang
, Y., Gu, Y., Chen, Z. and Han, Z.
Homo sapiens cDNA GK- clones
Unpublished (2000)
Contact: Zequang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919(ex.45)
Fax: 86-21-50801922
Email: hanzq@chac.sh.cn
This clone is available at CHGC in Shanghai.
```

```
Location/Qualifiers
1. 628
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="GKCB04"
/clone_lib="GKC"
/tissue_type="hepatocellular carcinoma"
/dev_stage="Adult"
/lab_host="SOLR"
/note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
XhoI"
BASE COUNT      164 a   127 c   160 g   176 t   1 others
ORIGIN

Query Match      43.4%; Score 322.8; DB 10; Length 628;
Best Local Similarity 77.6%; Pred. NO. 4.7e-78;
Matches 416; Conservative 0; Mismatches 117; Indels 3; Gaps 2;

Qy 81 atggacaaatagtaactcgtttgtctgtattctctgtctacttctctagctctcaaaagg 140
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 15 ACAGCACCCAAATGTAAGTCTGCCTTATGTCTCCAGCCNAITGCAAAAGGTCTCAAAAG 74

Qy 141 tctatggcctgtgttcactctgtattttggcctgaaaccccatagtgctgcatgcat 200
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 75 TCTATGGTCTGTGTTCACCGTGTATTTGGCATGAATCCCATAGTGTGTTCATGAT 134

Qy 201 atgaagcagtgaaagagccctgattgatctctggagagagatttctgaaagagcattt 260
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 135 ATGAGCAGCTGAAGGAAGCCCTGATTGATAATGGAGAGGAGTTTCTGGAAGAGGCAAT 194

Qy 261 tccactgctgaaagagctaacagagatttgatgtgcatgtgcatgtccgtgtttcaca 320
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 195 CCCCAATATCTCAAGAATTAAGGACTTGGTAGGTGCACATATTTCTGTGTGACT 254

Qy 321 tctgtcttgaggatgggagagatggaacacagacttacagagctcctcggcagagct 380
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 255 TTGGTAACCTGGGTGAGGGGATGGAACACAGAGCCCTAAAAAGCTTCTC-AGCAGACT 313

Qy 381 tggccatcacatagctgcccagctgt--cagcttcctcttcttgcctggatcctcct 438
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 314 TAGCTATCTGATGGCTGGCCGAGTGTTCGACAGCTTTCCTCTTGGCTGTGAAATCTCC 373

Qy 439 cctagtttctctctctctgttaggaattgttttcagcaatggaagaaatggaagga 498
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 374 CAGTTCTGCCCTTTTITTTATTAGGAATCATTTCCAGCAATGCAAGAGATGGAAGGA 433

Qy 499 gatccggcttctcctcctacgcctgcggaatttgggagtggaagagagagcattga 558
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 434 GATCCGGCGTTTCTCCCTCACAACTTGGCGAATTTGGGATGGGAAGAGAGGAGATTGA 493

Qy 559 gaaccgtgttcaagaggaagccgcctgctgtgtgagagagttgagaaaaaaccaagg 614
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 494 GGACCGTGTTCAGAGGAAGCTCACTGCTTGTGGAGGAGTTGAGAAAACCAAGG 549

RESULT 3
AV695514      631 bp   mRNA      EST      25-SEP-2000
LOCUS      AV695514 GK Homo sapiens cDNA clone GKCBID04 5', mRNA sequence.
DEFINITION      AV695514
ACCESSION      AV695514
VERSION      AV695514.1 GI:10297377
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 631)
AUTHORS      Wu, T., Qian, B., Huang, Q., Huang, C., Kang, B., Gao, X., Xu, Z., Xiao, H.,
, Xu, X., Li, N., Peng, Y., Liu, F., Qu, J., Song, H., Cheng, Z., Zeng, L.,
Xu, S., Gu, W., Tu, Y., Jia, J., Fu, G., Ren, S., Zhong, M., Lu, G., Yang
, Y., Gu, Y., Chen, Z. and Han, Z.
Homo sapiens cDNA GK- clones
Unpublished (2000)
Contact: Zequang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919(ex.45)
Fax: 86-21-50801922
Email: hanzq@chac.sh.cn
This clone is available at CHGC in Shanghai.
```

```
JOURNAL Unpublished (2000)
COMMENT Contact: Zeguang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919(ex.45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.

FEATURES
source
1..631
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="GKCBID04"
/tissue_type="hepatocellular carcinoma"
/dev_stage="Adult"
/lab_host="SOLR"
/Note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
XhoI"

BASE COUNT 164 a 128 c 160 g 178 t 1 others
ORIGIN

Query Match 43.4%; Score 322.8; DB 10; Length 631;
Best Local Similarity 77.6%; Pred. No. 4.7e-78;
Matches 416; Conservative 0; Mismatches 117; Indels 3; Gaps 2;

Oy 81 atgacaaaatagtaacttggttgctgttatctctgtctacttctcctagctctcaagg 140
Db 16 ACGAGCACCAATGTAAGTCGCTTATGTCCTCCAGCCCAATTGCAAAAGGTTCTCAAAG 75

Oy 141 tctatgacctgttctcaactctgtatttggcctgaaccataagtggtctacatgat 200
Db 76 TCTATGGCTGTGTTCCAGCGTATTTGGCATGAATCCCATAGTGTTTCATGAT 135

Oy 201 atgaagcagtgaaagagccctgattgatcttgagagagatttcttggaagagcattt 260
Db 136 ATGAGGCAGTCAAGGAAGCCCTGATTGATAATGGAGAGGAGTTTCTGGAAGAGGCAAT 195

Oy 261 tccactggctgaagagctaaacagagatttgtagtgatgctgctgtcttcaagca 320
Db 196 CCCCAATATCTCAAGAAATTTACTAAAGGACTTGGTAGTGACATATTTCTGTCTAGCT 255

Oy 321 tctgtcttgggagtggaagagctggaacacagagacttacagagctcctcgggcagagct 380
Db 256 TTGGTAACCTGGGGTGAGGGGATGGAACACAGAGCCCTTAAAGCTTCTC-AGCAGAGCT 314

Oy 381 tggccatccacatggctgcccagtg--cagcttctcttcttgcctggatctcct 438
Db 315 TAGCCTATCTGCATGGCTGCCGAGTGTTGCAGACATTTCTTCCCTGTGTAATCTCC 374

Oy 439 cctagttcgtttctctctctgttaggaattgtttcagcaatggaagaaatggaagga 498
Db 375 CAGTTTCTGCCCTTTTATTAGGAATCATTTCCAGCAATGGAAGAGATGGAAGGA 434

Oy 499 gatccggcgtttctcctctcagctgcggaattttgggatgggagagagcattga 558
Db 435 CATCCGGGTTTCTCCCTCACAACCTTCGGAATTTTGGGATGGGAAGAGAGCATIGA 494

Oy 559 ggaacgtgttcaagaggaagcccgctgctgtgagaggttgagaaaaaccagg 614
Db 495 GGACCGTGTTCAGAGGAAGCTCACTGCTTGTGGAGAGTTGAGAAAAACCAAG 550

RESULT 4
LOCUS AV697749 592 bp mRNA EST 25-SEP-2000
DEFINITION AV697749 GK Homo sapiens cDNA clone GKCBIF10 5', mRNA sequence.
ACCESSION AV697749
VERSION AV697749.1 GI:10299612
KEYWORDS EST.
SOURCE human.
```

```
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 692)
AUTHORS Wu,T., Qian,B., Huang,Q., Kang,B., Gao,X., Xu,Z., Xiao,H.,
Xu,X., Li,N., Peng,Y., Liu,F., Ou,J., Song,H., Cheng,Z., Zeng,L.,
Xu,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu,G., Yang
Y., Gu,Y., Chen,Z. and Han,Z.
Homo sapiens CDNA GK- clones
Unpublished (2000)
JOURNAL Contact: Zeguang Han
COMMENT Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919(ex.45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.

FEATURES
source
1..692
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="GKCBIF10"
/tissue_type="hepatocellular carcinoma"
/dev_stage="Adult"
/lab_host="SOLR"
/Note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
XhoI"

BASE COUNT 182 a 144 c 172 g 194 t
ORIGIN

Query Match 43.4%; Score 322.8; DB 10; Length 692;
Best Local Similarity 77.6%; Pred. No. 4.8e-78;
Matches 416; Conservative 0; Mismatches 117; Indels 3; Gaps 2;

Oy 81 atgacaaaatagtaacttggttgctgttatctctgtctacttctcctagctctcaagg 140
Db 18 ACGAGCACCAATGTAAGTCGCTTATGTCCTCCAGCCCAATTGCAAAAGGTTCTCAAAG 77

Oy 141 tctatgacctgttctcaactctgtatttggcctgaaccataagtggtctcctgagat 200
Db 78 TCTATGGCTGTGTTCCAGCGTATTTGGCATGAATCCCATAGTGTTCTGTCTAGAT 137

Oy 201 atgaagcagtgaaagagccctgattgatcttgagagagatttcttggaagagcattt 260
Db 138 ATGAGGCAGTGAAGGAAGCCCTGATTGATAATGGAGAGGAGTTTCTTGAAGAGGCAAT 197

Oy 261 tccactggctgaagagctaaacagagatttgtagtgatgctgctgttctcaagca 320
Db 198 CCCCAATATCTCAAGAAATTTACTAAAGGACTTGGTAGTGACATATTTCTGTCTAGCT 257

Oy 321 tctgtcttgggagtggaagagctggaacacagagacttacagagctcctcgggcagagct 380
Db 258 TTGTAACCTGGGGTGAGGGGATGGAACACAGAGCCCTTAAAGCTTCTC-AGCAGAGCT 316

Oy 381 tggccatccacatggctgcccagtg--cagcttctcttcttgcctggatctcct 438
Db 317 TAGCCTATCTGCATGGCTGCCGAGTGTTGCAGACATTTCTTCCCTGTGTAATCTCC 376

Oy 439 cctagttcgtttctctctctgttaggaattgtttcagcaatggaagaaatggaagga 498
Db 377 CAGTTTCTGCCCTTTTATTAGGAATCATTTCCAGCAATGGAAGAGATGGAAGGA 436

Oy 499 gatccggcgtttctcctctcagctgcggaattttgggatgggagagagcattga 558
Db 437 GATCCGGGTTTCTCCCTCACAACCTTCGGAATTTTGGGATGGGAAGAGAGCATIGA 496

Oy 559 ggaacgtgttcaagaggaagcccgctgctgtgagaggttgagaaaaaccagg 614
Db 497 GGACCGTGTTCAGAGGAAGCTCACTGCTTGTGGAGAGTTGAGAAAAACCAAG 552
```

```

RESULT 5
AV697837
LOCUS      551 bp      mRNA      25-SEP-2000
DEFINITION GRK Homo sapiens cDNA clone GKCBCF05 5', mRNA sequence.
ACCESSION AV697837
VERSION    AV697837.1 GI:10299700
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 651)
AUTHORS   Wu,T., Qian,B., Huang,Q., Kang,B., Gao,X., Xu,Z., Xiao,H.,
          Xu,X., Li,N., Peng,Y., Liu,F., Ou,J., Song,H., Cheng,Z., Zeng,L.,
          Xu,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu,G., Yang
          Y., Gu,F., Chen,Z. and Han,Z.
Homo sapiens cDNA GK- clones
Location/Qualifiers
     1..651
     /organism="Homo sapiens"
     /db_xref="taxon:9606"
     /clone="GKCBCF05"
     /clone_lib="GKC"
     /tissue_type="hepatocellular carcinoma"
     /dev_stage="Adult"
     /lab_host="SOLR"
     /note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
XhoI"
BASE COUNT 170 a 129 c 161 g 185 t      6 others
ORIGIN

Query Match      43.3%; Score 321.8; DB 10; Length 651;
Best Local Similarity 77.4%; Pred. No. 9e-78;
Matches 415; Conservative 0; Mismatches 118; Indels 3; Gaps 2;

Qy 81 atgacacaaatagtaacttggttggtgttatctctgtctactcttcttagctctcaaaagg 140
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1  ACGAGCACCAATGTAAGTGTGCGCTTATGTTCTCCAGCCAATGCAAAAGGTTTCAAAAG 60

Qy 141 tctatggccctgtgttcacctctgtatttggccctgaaacccataagtgctgctgcatgat 200
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 61 TCATGTGTCCTGTTGTTACCGTGATTTTGGCATGAATCCCAATGAGTGGTGTTCATGAT 120

Qy 201 atgaagcagtgaaagagccctgattgatcttgagagagagtttctctgagagagcatatt 260
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 121 ATGAGGCAGTGAAGGAAGGCCCTGATTGATATGAGAGAGGATTTTCTGGAAGAGGCAAT 180

Qy 261 tccactggctgaaagagctaaacagagattgttgatgtgcatgtgcctcttttcagaca 320
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 181 CCCCAATATCTCAAGAATTAATAAGGACTTGGTAGTGACACATATTCTCTGTGTCAGCT 240

Qy 321 tctgtcttggggatggggagagatgaaacacagacactacagagctcctcgggcagagct 380
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 241 TTGGTAACTGGGGTGGAGGGGATGGAAACAGAGCCCTAAAAGAGTCTTC-AGCAGAGCT 299

Qy 381 tggcccatccacatggctgccagtg- -cagcttctcttcttcttgcctggagatcctct 438
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 300 TAGCCTATCTGCATGGCTGCCAGTGTTCAGACACTTCTCTCTGGCTGTGATCTCTCC 359

Qy 439 cctagtttctctcttcctgttaggaattgttttcagcaatgaaagaaatgcgaagg 498
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 360 CAGTTCTCGCCCTTTTTTTTATTAGGAATCAATCTCCAGCAATGAAAGAGATGTGAAGGA 419

```

Qy: 499 gatccgcgtttctccctcatgacgctcgggaattttggatgggggaagagacattga 558
 Db 420 GATCCGCGCTTCTCCCTCAACACTTCGCGAATTTGGGATGNGAAGAGGACATTTGA 479
 Qy 559 ggaccgtgttcaagaagaagccgctccttctgtggagagttgagaaaaacccaagg 614
 Db 480 GGACCGTGTTCAAGAGGAAGCTCACTGCCITGTGGAGGAGTTGAGAAAAACCAAGG 535

RESULT 6
 AV696436 AV696436 675 bp mRNA EST 25-SEP-2000
 LOCUS AV696436 GKC Homo sapiens cDNA clone GKCBAJ05 5', mRNA sequence.
 DEFINITION AV696436
 ACCESSION AV696436
 VERSION AV696436.1 GI:10298299
 KEYWORDS Est.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 REFERENCE 1 (bases 1 to 675)
 AUTHORS Wu, T., Qian, B., Huang, Q., Huang, C., Kang, B., Gao, X., Xu, Z., Xiao, H.,
 Xu, X., Li, N., Peng, Y., Liu, F., Ou, J., Song, H., Cheng, Z., Zeng, L.,
 Xu, S., Gu, W., Tu, Y., Jia, J., Fu, G., Ren, S., Zhong, M., Lu, G., Yang,
 Y., Gu, Y., Chen, Z. and Han, Z.
 TITLE Homo sapiens cDNA GK- clones
 JOURNAL Unpublished (2000)
 COMMENT Contact: Zeqiang Han
 Chinese National Human Genome Center at Shanghai
 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
 201203, P. R. China
 Tel: 86-21-50801919(ex.45)
 Fax: 86-21-50801922
 Email: hanzq@chgc.sh.cn
 This clone is available at CHGC in Shanghai.

FEATURES
 source
 1..675
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="GKCBAJ05"
 /clone_lib="GKC"
 /tissue_type="hepatocellular carcinoma"
 /dev_stage="Adult"
 /lab_host="SOLR"
 /note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
 XhoI"
 BASE COUNT 178 a 136 c 162 g 198 t 1 others
 ORIGIN

Query Matchbase 43.3%; Score 321.8; DB 10; Length 675;
 Best Local Similarity 77.4%; Pred. No. 9e-78;
 Matches 415; Conservative 0; Mismatches 118; Indels 3; Gaps 2

Qy 81 atggacaaaatagtaacttcgttgcgtttatctctgtctactttctctagctctcaagg 140
 Db 1 ACGAGCCAAATGTAAGTCTGCCTATGTTCTCCAGCCAAATGCAAGGTTCTCAAAAG 60
 Qy 141 tctatggccctgtgttcaactctgtattttggctgaaacccatagtgctgcatgat 200
 Db 51 TCTATGTCCTGTGTTTCACCGTGATTTTGCAATGAATCCCATATGCTGTTTCATGAT 120
 Qy 201 atgaagcagtgaaaggaagccctgatgatcttggaggagatcttctggaagaggcatt 260
 Db 121 ATGAGGCAGTGAAGGAAGCCCTGATTGATATGATGGAGAGAGGATTTTCTGGAAGAGCAAT 180
 Qy 251 tcccactgctgaaagaagcctaacagagatttggtagtgctgcatgtccctgttcagca 320
 Db 181 CCCCAAATATCTCAAGAATTAATAAGGACTTGGTAGGTGCACATATTCTGTGTGACGT 240
 Qy 321 tctgtcttgggagtgaggagatgaaacagagacttacagagctcctcggggcagagct 380

[illegible]

Db	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385	386	387	388	389	390	391	392	393	394	395	396	397	398	399	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420	421	422	423	424	425	426	427	428	429	430	431	432	433	434	435	436	437	438	439	440	441	442	443	444	445	446	447	448	449	450	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466	467	468	469	470	471	472	473	474	475	476	477	478	479	480	481	482	483	484	485	486	487	488	489	490	491	492	493	494	495	496	497	498	499	500	501	502	503	504	505	506	507	508	509	510	511	512	513	514	515	516	517	518	519	520	521	522	523	524	525	526	527	528	529	530	531	532	533	534	535	536	537	538	539	540	541	542	543	544	545	546	547	548	549	550	551	552	553	554	555	556	557	558	559	560	561	562	563	564	565	566	567	568	569	570	571	572	573	574	575	576	577	578	579	580	581	582	583	584	585	586	587	588	589	590	591	592	593	594	595	596	597	598	599	600	601	602	603	604	605	606	607	608	609	610	611	612	613	614	615	616	617	618	619	620	621	622	623	624	625	626	627	628	629	630	631</
----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-------

Query Match	43.3%	Score 321.8;	DB 10;	Length 677;
Best Local Similarity	77.4%;	Pred. No. 9e-78;		
Matches	415;	Conservative	0;	Mismatches 118;
				Indels 3;
				Gaps 2

```
Oy 81 atggacaaaatagtaactcgttctgctgttatctctgtctactcttctctagctctcaagg 140
Db 2 ACAGCACCACCAATGTAAGTCTGCTATGTTCTCCAGGCAATTCACAAAGGTTCTCAAAAG 61

Oy 141 tctatggccctgttctcactctgtattttggcctgaaccatagtgctgctgcacgat 200
Db 62 TCTATGGTCTGTGTTCCACCGTGATTTTGGCATGATCCCATAGGTGTTTCAAGAT 121

Oy 201 atgaagcagtggaaggagccctgattgatcttggaggagaggtttcttggaaagagcattt 260
Db 122 ATGAGGACAGTGAAGGAGCCCTCAITGATAATGGAGAGAGTTCCTGGAAGAGGCAATT 181

Oy 261 tccactggctgaagacgctaagcagagatttggtaagtgatcagatgcctcttttcagca 320
Db 182 CCCATAATCTCAAGAAATTAATAAGGACTTGGTAGGTGCACATATTTCTGTGTCAGCT 241

Oy 321 tctgttctgggagtgaggagtggaacacagagacttacagagctcctcgggagcagct 380
Db 242 TTGTAATCTGGGTGAGGGGATGGAACACAGAGCCCTAAAGAGCTTCTC-AGCAGAGCT 300

Oy 381 tggccatcacatgctgctcccaagt--cagcttctcttcttctgctggaatctcct 438
Db 301 TAGCCTATCTGATGCTGCCAGTGTGCAGCATTCTTCTTGGCTGTGAAATCTCCTC 360

Oy 439 cctagttctcttctctctgttagaatgttttcagcaatggaagaaatggaagga 498
Db 361 CAGTTCTGCCCTTTTATTAGGAATCATTTCCAGCAATGGAAGAGATGGAAGGA 420

Oy 499 gatccggcgtttctcctcatgacgctgcggaaatttggatgggaaagagagcaattga 558
Db 421 GATCCGCGGCTTTCTCCCTCACACCTTGGGGAATTTGGAGTGGNCAAGAGGAGCATTC 480

Oy 559 ggaccgtgttcaagaggaagccgcctgtgcttggaggaggttgagaaacccaagg 614
Db 481 GGACCGGTGTTCAAGAGGAAGCTCACTGCCTTGTGGAGGAGTTGAGAAAAACCAAGS 536

RESULT 9
AV696639 580 bp mRNA EST 25-SEP-2000
LOCUS AV696639 GK Homo sapiens cDNA clone GKCBIE09 5', mRNA sequence.
DEFINITION AV696639
ACCESSION AV696639
VERSION AV696639.1 GI:10298502
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHOR Wu,T., Qian,B., Huang,Q., Kang,B., Gao,X., Xu,Z., Xiao,H.,
Xu,X., Li,N., Peng,Y., Liu,F., Ou,J., Song,H., Cheng,Z., Zeng,L.,
Xu,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu,G., Yang
,Y., Gu,Y., Chen,Z. and Han,Z.
Homo sapiens cDNA GK- clones
Unpublished (2000)
Contact: Zequan Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919(ex.45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.
Location/Qualifiers
1. 680
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="GKCBIE09"
/clone_lib="GKC"
/tissue_type="hepatocellular carcinoma"
/dev_stage="Adult"
/lab_host="SOLR"
/note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
```

```
BASE COUNT 177 a 141 c 167 g 192 t 3 others
ORIGIN
Query Match 42.7% Score 317: DB 10: Length 680:
Best Local Similarity 76.9% Pred. No. 1.9e-76:
Matches 412: Conservative 0: Mismatches 121: Indels 3: Gaps 2:

Oy 81 atggacaaaatagtaactcgttctgctgttatctctgtctactcttctctagctctcaagg 140
Db 11 ACGAGCACCACCAATGTAAGTCTGCTATGTTCTCCAGGCAATTCACAAAGGTTCTCAAAAG 70

Oy 141 tctatggccctgttctcactctgtattttggcctgaaccatagtgctgctgcacgat 200
Db 71 TCTATGGTCTGTGTTCCACCGTGATTTTGGCATGATCCCATAGGTGTTTCAAGAT 130

Oy 201 atgaagcagtggaaggagccctgattgatcttggaggagaggtttcttggaaagagcattt 260
Db 131 ATGAGGCACTGAGGAGCCCTGATTGATAATGGAGAGGAGTTCCTGGAAGAGGCAATT 190

Oy 261 tccactggctgaagacgctaagcagagatttggtaagtgatcagatgcctcttttcagca 320
Db 191 CCCATAATCTCAAGAAATTAATAAGGACTTGGTAGGTGCACATATTTCTGTGTCAGCT 250

Oy 321 tctgttctgggagtgaggagtggaacacagagacttacagagctcctcgggagcagct 380
Db 251 TTGTAATCTGGGTGAGGGGATGGAACACAGAGCCCTAAAGAGCTTCTC-AGCAGAGCT 309

Oy 381 tggccatcacatgctgctcccaagt--cagcttctcttcttctgctggaatctcct 438
Db 310 TAGCCTATCTGATGCTGCCAGTGTGCAGCATTCTTCTTGGCTGTGAAATCTCCTC 369

Oy 439 cctagttctcttctctctgttagaatgttttcagcaatggaagaaatggaagga 498
Db 370 CAGTTCTGCCCTTTTATTAGGAATCATTTCCAGCAATGGAAGAGATGGAAGGA 429

Oy 499 gatccggcgtttctcctcatgacgctgcggaaatttggatgggaaagagagcaattga 558
Db 430 GATCCGCGGCTTTCTCCCTCACACCTTGGGGAATTTGGAGTGGNCAAGAGGAGCATTC 489

Oy 559 ggaccgtgttcaagaggaagccgcctgtgcttggaggaggttgagaaacccaagg 614
Db 490 GGACCGGTGTTCAAGAGGAAGCTCACTGCCTTGTGGAGGAGTTGAGAAAAACCTCAAG 545

RESULT 10
AV697579 622 bp mRNA EST 25-SEP-2000
LOCUS AV697579 GK Homo sapiens cDNA clone GKCBJB12 5', mRNA sequence.
DEFINITION AV697579
ACCESSION AV697579
VERSION AV697579.1 GI:10299442
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHOR Wu,T., Qian,B., Huang,Q., Kang,B., Gao,X., Xu,Z., Xiao,H.,
Xu,X., Li,N., Peng,Y., Liu,F., Ou,J., Song,H., Cheng,Z., Zeng,L.,
Xu,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu,G., Yang
,Y., Gu,Y., Chen,Z. and Han,Z.
Homo sapiens cDNA GK- clones
Unpublished (2000)
Contact: Zequan Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919(ex.45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.
Location/Qualifiers
```

```
source          1. 622
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="GKCBJ12"
/clone_lib="GKC"
/tissue_type="hepatocellular carcinoma"
/dev_stage="Adult"
/lab_host="SOLR"
/Note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
XhoI"
BASE COUNT    165 a 126 c 156 g 175 t
ORIGIN

Query Match      42.4%; Score 314.8; DB 10; Length 622;
Best Local Similarity 76.7%; Pred. No. 7.5e-76;
Matches 411; Conservative 0; Mismatches 122; Indels 3; Gaps 2:

QY 81 atggacaaataaactctgttctgtctatctctgtctacttctcctagctctcaaaag 140
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1 ACCAGCACCAATGTAGCTGCGCTTATGTCTCTACAGCCAAATTCGAAGGTTCTCAAAAG 60

QY 141 tctatggccctgttctcctctgtattttggcctgaaacccatagtgctctcattgat 200
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 61 TCTATGCTCTGTGTTTACCGCTGTATTTGGCATGATCCCATAGTGGTGTTCATGGAT 120

QY 201 ataaacagtgaaagacccctgattgattcttggagagagatttcttggaaagagcattt 260
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 121 ATGAGGCGAGTGAACGAGCCCTGATTGATTAATGGAGGAGGATTCCTCGAAGAGGCAAT 180

QY 261 tccactggctgaaagagctaaacagaggtttgtgaggtgtgctgtcctgtttcagca 320
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 181 CCCCAATATCTCAAGAAATTTACTAAAGGACTTGTAGGTGCACATATTTCTGTGTCAGCT 240

QY 321 tctgtcttggagtgaggagatggaacacagagacttacagagctcctcgggcagagct 380
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 241 TTGGTAAGTGGAGTGGAGCGGATGGAACACAGAGCCCTTAAAGAGCTTCTC-AGCAGAGCT 299

QY 381 tggcccatcacatggctgcctcagtg--cagcttctcttcttctgcctggagctcctc 438
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 300 TAGCCATCTGCATGGCTGCCGAGTGTTCAGCAGCATTTCTTCTTGGCTGTGAATTTCTCC 359

QY 439 cctagttctcttctctctctgttaggaattgttttcagcaatggaagaaatgggaagga 498
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 360 CAGTTTCTGCCCTCTTTTATTTAGGAATCATTTCCAGCAATGGACAGATGGGAAGGA 419

QY 499 gatcccgctttctcctcatgagctgcggaattttggatagggaagagagcattga 558
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 420 GATCCGCGCTTCTCCCTCACACCTTTCGGAATTTTGGATGGGAAGGAGGACATTGA 479

QY 559 ggaccgtgttcaagagaaagccgcctgtgtgaggagattgaagaaacccaag 514
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 480 GGACCGTGTTCAGAGAGAGCTCACTGCCTTGTGGAGGAGTTCAGAAAAAACCAAGG 535

RESULT 11
AV646984
LOCUS          747 bp mRNA EST 07-SEP-2000
DEFINITION    GLC Homo sapiens cDNA clone GLCATD11 3', mRNA sequence.
ACCESSION    AV646984
VERSION      AV646984.1 GI:9867998
KEYWORDS     EST.
SOURCE       human.
ORGANISM     Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1 (bases 1 to 747)
AUTHORS      Qian,B., Wu,T., Huang,Q., Huang,C., Kang,B., Gao,X., Xu,Z., Xiao,H.,
Zeng,L., Xu,S., Gu.W., Tu,Y., Jia,J., Fu.G., Ren.S., Zhong,M., Lu
G., Yang,Y., Gu.Y., Chen.Z. and Han.Z.
Homo sapiens cDNA clone
Unpublished (2000)

Contact: Zeguqiang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919(ex.45)
Fax: 86-21-50801922
Email: hanzq@chgc.sh.cn
This clone is available at CHGC in Shanghai.

FEATURES
Location/Qualifiers
1..747
Source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="GLCATD11"
/clone_lib="GLC"
/tissue_type="corresponding non cancerous liver tissue"
/dev_stage="Adult"
/lab_host="SOLR"
/Note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
XhoI"
BASE COUNT    212 a 141 c 186 g 206 t 2 others
ORIGIN

Query Match      39.0%; Score 289.6; DB 10; Length 747;
Best Local Similarity 98.6%; Pred. No. 6.5e-69;
Matches 292; Conservative 0; Mismatches 4; Indels 0; Gaps 0:

QY 1 tcagaaatatttgaagcctgtgtgctgaataaaagacatacaaatgaaatatca 60
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 192 TCAGAAATATTTGAAGCCTGTGTGGCTGAATAAAAGCATACAAATACAAATGAAATATCA 251

QY 61 tctaaatcagccttagcaaatgacaaaatagtaacttgcctgttgccttctctctct 120
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 252 TCTAAATCAGGCTTAGCAATAGGACAAATAGTAACCTTGTGTTGCTGTCTGTCTGTCT 311

QY 121 actttctgactctcaaaaggtctatggccctgtgttccactctgtattttggcctgaaacc 180
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 312 ACTTTCCTAGCTCTAAAGGTCTATGGCCCTGTGTCTGTATTTTGGCTGAAACC 371

QY 181 catagtggctgcctcatgatatgaagcagtgaaagacccctgattgacttggagaga 240
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 372 CATAGTGGTGTGCTGATATGAAGCAGTGAAGGAGCCCTGATTTGATCTTTGGAGAGA 431

QY 241 gttttctgaaagaggcattttccctcactgctgaaagactaacagagatttgta 296
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 432 GTTTCGGAAGAGGCAATTTTCCCACTGGCTGAAGAGCTACAGAGGATTGGAA 487

RESULT 12
AV656485
LOCUS          646 bp mRNA EST 25-SEP-2000
DEFINITION    GKC Homo sapiens cDNA clone GKCBJA12 5', mRNA sequence.
ACCESSION    AV656485
VERSION      AV656485.1 GI:10298348
KEYWORDS     EST.
SOURCE       human.
ORGANISM     Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1 (bases 1 to 646)
AUTHORS      Wu,T., Qian,B., Huang,Q., Huang,C., Kang,B., Gao,X., Xu,Z., Xiao,H.,
Xu,X., Li.N., Peng.Y., Liu,F., Qu,J., Song,H., Cheng,Z., Zeng,L.,
Xu.S., Gu.W., Tu.Y., Jia.J., Fu.G., Ren.S., Zhong.M., Lu.G., Yang
Y., Gu.Y., Chen.Z. and Han.Z.
Homo sapiens cDNA GK- clones
Unpublished (2000)
Contact: Zeguqiang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919(ex.45)
Fax: 86-21-50801922
Email: hanzq@chgc.sh.cn
```

This clone is available at CHGC in Shanghai.

FEATURES

source

Contact: Robert Strausberg, Ph.D.
Email: cchapbs-r@mail.nih.gov
Tissue Procurement: CLONETECH Laboratories, Inc.
CDNA Library Preparation: CLONETECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LILCM1628 row: b column: 11
High quality sequence stop: 718.
Location/Qualifiers
1..721
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4766890"
/clone_lib="NIH_MGC_76"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: Liver; Vector: pDNR-LIB (Clontech); Site_1: SfiI (ggccattcgccc); Site_2: SfiI (ggccattcgccc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCGCGGCGGCATATG-dt(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.85 kb (range 1.0-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH_MGC Library."

BASE COUNT
ORIGIN

176 a 147 c 195 g 203 t

Query Match 36.4%; Score 270.4; DB 11; Length 721;
Best Local Similarity 99.3%; Pred. No. 1.2e-63;
Matches 282; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

Qy 461 ttaggcaattgtttcagcaatggaagaaatggaaggagatccggcggtttccctcaltg 520
Db 355 TTGTGAATTTGTTTTCAGCAATGGAAGAAATGGAAGAGATCCGGCGTTTCTCCCTCATG 414
Qy 521 acgtctcggaatttggatggggaaggagcattgagaccgtgttcaagagaagcc 580
Db 415 ACCTGCGGAAATTTTGGGATGGGGAAGAGAGCATTTGAGGACCGTGTTCAGAGAGAGCC 474
Qy 581 cgtctgcttctggaagggttgagaaaaaaccaagggtgggtgacctactccatcactg 640
Db 475 CGCTGCTTGTGAGGAGTGTGAGAAAAACCAAGGTTGGTGTACCTACTATCATCATG 534
Qy 641 accttactggactactcttcttactgacattcttgaacatttcagagggtggccat 700
Db 535 ACCTTACTGGACTACTATCTTCTTACTGACATTTCTTGAACATTTTCAGGGGTGGCCAT 594
Qy 701 atcttctattatgagt-ctgggtgttagctcatgtgaagcgagg 743
Db 595 ATCTTTCAITATGAGTCCTGGTTGTGTAGTCTATGTGTAAGCGGGG 638

RESULT 14

AV595838
LOCUS AV595838 530 bp mRNA EST 25-SEP-2000
DEFINITION AV595838 GK C Homo sapiens cDNA clone GKCBEE06 5', mRNA sequence.
ACCESSION AV595838
VERSION AV595838.1 GI:10297701
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 530)
AUTHORS Wu, J., Qian, B., Huang, Q., Kang, B., Gao, X., Xu, Z., Xiao, H., Xu, X., Li, N., Peng, Y., Liu, F., Ou, J., Song, H., Cheng, Z., Zeng, L., Xu, S., Gu, W., Tu, Y., Jia, J., Fu, G., Ren, S., Zhong, M., Lu, G., Yang, Y., Gu, Y., Chen, Z. and Han, Z.

BASE COUNT 172 a 130 c 161 g 183 t

ORIGIN

Query Match 37.7%; Score 280.2; DB 10; Length 646;
Best Local Similarity 72.6%; Pred. No. 2.5e-66;
Matches 390; Conservative 0; Mismatches 143; Indels 4; Gaps 2;
Qy 81 atggacaaaatagtaacttggttgtgttatctctgtctacttctcagctctcaagg 140
Db 2? ACGAGCAACCAATTAAGTCTGCCCTTATGTTCTTACAGCCAAATTCGAAAGGTTCTCAAAAG 81
Qy 141 tctatggccctgtttcactctgtatttggcctgaaacccatagtggtctcctcagat 200
Db 82 TCTATGCTCTGTGTTCCACCTGTATTTGGCATGAATTCATAGAGGTGTTTATGGAT 141
Qy 201 atgaagcagtgaaaggagccctgattgatcttggaggagagatttctgaaggagcatt 260
Db 142 ATGAGGCAGTGAACGAAGCCCTGATTGATGATGAGAGTAGGTCITTTGGAAGAGGCAATT 201
Qy 261 tccactggctgaaagagtcacagagatttggtagtgatgctgctctcttcacga 320
Db 202 ACCCAATATCTCAAGAAATTAATTAAGGACTTGGTAGGTGCACATATATCTGTCTAGCT 261
Qy 321 tctgtcttggggatgggagatggaaacagagacttacagagctcctcggagcagct 380
Db 262 TTGTAACCTGGAGTGGGGCGATGGAAACAGAGAGCCCTAAAGAGCTAGTC-ATCAGATCT 320
Qy 381 tggccatcacatggctgccagtgctcagcttctctt-cttgcctggagctccc 437
Db 321 TATCTATCTGCAATGAGTGGAGTGTGTGATCATCTTCTACCTTGGCTGTGNAATATCTC 380
Qy 438 tctagttctcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttct 497
Db 381 CCAGATTCTGCCCTGAAATTTTATTAGGAATGATTTCCAGCAATGGACAGACATGGAAG 440
Qy 498 agatccggcgtttctccctcagcgtcggaatttggatggggaagagagcattg 557
Db 441 AGATCCGGGCTTCTCCCTCACACCTTGGCGGAATTTGGGATGGGAGAGGAGCATIG 500
Qy 558 aggaccgtgttcaagagagcccgctgcttctgtgagaggttgagaaaaaaccaagg 614
Db 501 AGGACCGTTTTCAGAGAGAGCTCACTGCTTGTGCGGAGTTGCGAAACCAAGG 557

RESULT 13

BG617892
LOCUS BG617892 721 bp mRNA EST 18-APR-2001
DEFINITION 602645541F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4766890 5', mRNA sequence.
ACCESSION BG617892
VERSION BG617892.1 GI:13669263
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 721)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>,
TITLE National Institutes of Health, Mammalian Gene Collection: (MGC)
JOURNAL Unpublished (1999)

<p>TITLE Homo sapiens cDNA GK- clones JOURNAL Unpublished (2000) COMMENT Contact: Zeguang Han Chinese National Human Genome Center at Shanghai 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai 201203, P. R. China Tel: 86-21-50801919(ex.45) Fax: 86-21-50801922 Email: hanzg@chgc.sh.cn This clone is available at CHGC in Shanghai.</p>	
<p>FEATURES</p>	<p>Location/Qualifiers</p> <p>1..530</p> <p>/organism="Homo sapiens"</p> <p>/db_xref="taxon:9606"</p> <p>/clone="GKBBE06"</p> <p>/clone_lib="GKC"</p> <p>/tissue_type="hepatocellular carcinoma"</p> <p>/dev_stage="Adult"</p> <p>/lab_host="SOLR"</p> <p>/note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2: XhoI"</p>
<p>BASE COUNT</p>	<p>124 a 108 c 147 g 151 t</p>
<p>Query Match 36.0% Score 267.4; DB 10: Length 530; Best Local Similarity 74.7% Pred. No. 8e-63; Matches 392: Conservative 0; Mismatches 121; Indels 12; Gaps</p>	
<p>Oy 81 atggacaaaatagtaactctgttgcgtgtatctctgtctacttctctagctctcaagg 140</p>	<p>Db 15 acgagcaccatgtaagtctgccttatgttctctccgcccaattgcacaaagggtctcaaaag 75</p>
<p>Oy 141 tctatggcctgtgttcaactctgtatttggcctgaaacccatagtgctcctcatgcat 200</p>	<p>Db 76 tctatgtctctgtgttccacgtgtatgttggcatgaaatcccatagtggtgtttcatggat 135</p>
<p>Oy 201 atgaagcagtgaaagagccctgtatgtcttgagagagatttcttggaagagcattt 260</p>	<p>Db 136 atgaggcagtgaaagagccctgattgataatgagagaggagtgttcttggaagagcattt 195</p>
<p>Oy 261 tcccactgtctgaaagagctaacagagatttggataggtgtgcatgtgcctgtttcagca 320</p>	<p>Db 196 ccccaaatatctcaagaataattactaaagagcttggtaggtgcacataatttctgtgtcagct 255</p>
<p>Oy 321 tctgtcttggagatgggagagatgaaacacagacttaccagactcctcggaacact 380</p>	<p>Db 256 ttggttaactgggggtgagggggatgaaacacagacccta----gccttctcagcagagct 310</p>
<p>Oy 381 tggccatccacatggctgccagtg--cagcttctcttcttctgtcctggatctccct 438</p>	<p>Db 311 tagcctatctgcattggctgcccagtggttgcagcactttcttctgtgtgaaattctcc 370</p>
<p>Oy 439 cctagtttcgtttctcttctctgttaggaattgttttcagcaatgaaagaatggaaagga 498</p>	<p>Db 371 cagttttctgccccctttttttatttaggaataatttcagcaatgg---gagatggaaagga 427</p>
<p>Oy 499 gatccggcgtttctccctcatgacgctgcggaatttggatggggaagcgagcattga 558</p>	<p>Db 428 gatccggcgtttctccctcacacacttgcggaaatttgggatcgsgaagagcagcaatga 487</p>
<p>Oy 559 ggaacgtgttcaagagaaagcccgctccttgttggaagatgag 603</p>	<p>Db 488 ggaacgtgttcaagagg--gctactgccttctggaagaggtttag 530</p>

RESULT	15
AV650643	
LOCUS	AV650643 395 bp mRNA
DEFINITION	AV650643 GLC Homo sapiens cDNA clone GCGCG12 3', mRNA sequence
ACCESSION	AV650643
VERSION	AV650643.1 GI:9881557
KEYWORDS	EST.

SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:
Mammalia: Eutheria: Primates: Catarrhini; Hominidae; Homo.
1 (bases 1 to 395)
AUTHORS Qian,B., Wu,T., Huang,Q., Huang,C., Kang,B., Gao,X., Xu,Z., Xiao,H.,
, Xu,X., Li,N., Peng,Y., Liu,F., Qu,J., Song,H., Cheng,Z., Qu,J.,
Zeng,L., Xu,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu
,G., Yang,Y., Gu,Y., Chen,Z. and Han,Z.
TITLE Homo sapiens cDNA clone
JOURNAL Unpublished (2000)
COMMENT Contact: Zequang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919(ex.45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.
FEATURES Location/Qualifiers
source
1..395
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="GLCGJ12"
/clone_lib="GLC"
/tissue_type="corresponding non cancerous liver tissue"
/dev_stage="Adult"
/lab_host="SOLR"
/note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
XhoI"
BASE COUNT 108 a 71 c 108 g 108 t
ORIGIN
Query Match 34.1%; Score 253.2; DB 10; Length 395;
Best Local Similarity 98.8%; Pred. No. 6.2e-59;
Matches 255; Conservative 0; Mismatches 3; Indels 0; Gaps 0
Oy 39 tacaataacaatgaaatatcatgctaataatcaggcttagcaaatgacacaaatagtaact 98
|||||
Db 1 TACAATACATGAANAATATCATGCTAATACAGCTTAGCAAAATGCAAAATAGTAAC 60
Oy 99 tctgttggctgttatctctgtctacttctcctagctctcaaggctcattgccccctgttcca 158
|||||
Db 61 TCGTITGCTGTATCTCTGCTACTTCTCTAGCTCTCAAGGCTCATGCGCCCTGTGTCCA 120
Oy 159 cctctattttggcctgaaccatagtggtgctgcattgattgaagcagtgaaaggaag 218
|||||
Db 121 CTCTGTATTTTGGCTTGAACCCATAGTGGTCTGCTCATGATATGAAGCAGTGAAGGAAG 180
Oy 219 cccctgattgattctggagagaggttttcttggaaaggcattttcccaactggctgaaagag 278
|||||
Db 181 CCCGTATTGATCTCGAGAGGAGTTTTCTTGGAAGAGGCATTTTCCCACTGGCTGAAGAG 240
Oy 279 ctacacagagatttgga 296
|||||
Db 241 CTACAGAGGATTGGAA 258

Search completed: April 19, 2002, 08:07:49
Job time: 2759 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 19, 2002, 08:10:30 : Search time 130.44 Seconds
(without alignments)
1290.043 Million cell updates/sec

Title: US-09-763-292-2
Perfect score: 743
Sequence: 1 tcagaaatattgaagcctg.....ttaagctcatgtgaacgagg 743

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 11328999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA.*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PTUS_COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed.
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Desc: Iption
1	163	21.9	1419	2	US-08-194-981E-4
2	163	21.9	1591	2	US-08-194-981E-3
3	163	21.9	1852	1	US-08-201-118-10
4	163	21.9	1852	2	US-08-238-821B-10
5	163	21.9	1852	5	PCT-US95-05744-10
6	163	21.9	1854	1	US-08-201-118-4
7	163	21.9	1854	2	US-08-238-821B-4
8	163	21.9	1854	5	PCT-US95-05744-4
9	153.4	20.6	1892	1	US-08-201-118-14
10	153.4	20.6	1892	2	US-08-238-821B-14
11	153.4	20.6	1892	5	PCT-US95-05744-14
12	149.6	20.1	1746	1	US-08-201-118-2
13	149.6	20.1	1746	2	US-08-238-821B-2
14	149.6	20.1	1746	5	PCT-US95-05744-2
15	144	19.4	200	1	US-08-716-459-13
16	140.2	18.9	2009	1	US-08-201-118-6
17	140.2	18.9	2009	2	US-08-238-821B-6
18	140.2	18.9	2009	5	PCT-US95-05744-6
19	140.2	18.9	2258	1	US-08-201-118-12
20	140.2	18.9	2258	2	US-08-238-821B-12
21	140.2	18.9	2258	5	PCT-US95-05744-12
22	134.2	18.1	1829	1	US-08-201-118-8
23	134.2	18.1	1829	2	US-08-238-821B-8
24	134.2	18.1	1829	5	PCT-US95-05744-8
25	89.4	12.0	100	1	US-08-716-459-12
26	79.2	10.7	1419	2	US-08-194-981E-2
27	79.2	10.7	1482	2	US-08-194-981E-1

28	63.6	8.6	7216	2	US-08-750-703-3	Sequence 3, Appli
29	63.6	8.6	8779	2	US-08-750-703-4	Sequence 4, Appli
30	59.8	8.0	1737	2	US-08-750-703-2	Sequence 2, Appli
31	59.8	8.0	1740	2	US-08-750-703-1	Sequence 1, Appli
32	44.2	5.9	50	1	US-08-716-459-11	Sequence 11, Appli
33	41	5.5	1568	2	US-08-145-658D-20	Sequence 20, Appli
34	39.4	5.3	153	6	5508199-3	Patent No. 5508199
35	39.4	5.3	1566	2	US-08-145-658D-13	Sequence 13, Appli
36	39.4	5.3	1566	2	US-08-145-658D-22	Sequence 22, Appli
37	39.4	5.3	1569	2	US-08-145-658D-23	Sequence 23, Appli
38	39.4	5.3	1571	2	US-08-145-658D-21	Sequence 21, Appli
39	39	5.2	1501	2	US-08-145-658D-24	Sequence 24, Appli
40	33.4	4.5	1509	1	US-08-115-052-1	Sequence 1, Appli
41	30.8	4.1	7218	1	US-08-232-463-14	Sequence 14, Appli
42	30	4.0	1317	1	US-08-153-848-45	Sequence 45, Appli
43	30	4.0	1317	3	US-09-299-843A-45	Sequence 45, Appli
44	30	4.0	1317	5	PCT-US93-11153-45	Sequence 45, Appli
45	30	4.0	7218	1	US-08-232-463-14	Sequence 14, Appli

ALIGNMENTS

RESULT 1
US-08-194-981E-4
: Sequence 4, Application US/08194981E
: Patent No. 5886157
: GENERAL INFORMATION:
: APPLICANT: GUENGERICH, F. Peter
: APPLICANT: GUO, Zuyu
: APPLICANT: SANDHU, Punam
: APPLICANT: GILLIAM, Elizabeth M. J.
: TITLE OF INVENTION: EXPRESSION AND PURIFICATION OF
: TITLE OF INVENTION: HUMAN
: TITLE OF INVENTION: CYTOCHROME P450
: NUMBER OF SEQUENCES: 68
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: NEEDLE & ROSENBERG, P.C.
: STREET: Suite 1200, 127 Peachtree Street, NE
: CITY: Atlanta
: STATE: Georgia
: COUNTRY: USA
: ZIP: 30303-1811
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/194,981E
: FILING DATE: February 10, 1994
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Elizabeth Selby
: REGISTRATION NUMBER: 38,298
: REFERENCE/DOCKET NUMBER: 22000.0022
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (404) 688-0770
: TELEFAX: (404) 688-9880
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1419 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: Single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: HYPOTHETICAL: NO
: ANTI-SENSE: NO
US-08-194-981E-4

Query Match 21.9% Score 163; DB 2: Length 1419;
Best Local Similarity 97.1%; Pred. No. 1.4e-42;

RESULT 5

```

: GENERAL INFORMATION:
: APPLICANT: GOLDSTEIN, Joyce A.
: APPLICANT: ROMKES-SPARKS, Marjorie
: TITLE OF INVENTION: CLONING AND EXPRESSION OF COMPLEMENTARY
: TITLE OF INVENTION: DNAs FOR MULTIPLE MEMBERS OF THE HUMAN CYTOCHROME P450 2C
: TITLE OF INVENTION: SUBFAMILY
: NUMBER OF SEQUENCES: 44
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Townsend and Townsend Kourile and Crew
: STREET: 379 Lytton Avenue
: CITY: Palo Alto
: STATE: California
: COUNTRY: US
: ZIP: 94301
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/201,118
: FILING DATE: 22-FEB-1994
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/864,962
: FILING DATE: 09-APR-1992
: ATTORNEY/AGENT INFORMATION:
: NAME: Liebeschuetz, Joe
: REGISTRATION NUMBER: 37,505
: REFERENCE/DOCKET NUMBER: 15280-192-1
: TELEPHONE: (415) 326-2400
: TELEFAX: (415) 326-2422
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1854 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: US-08-201-118-4

Query Match 21.9%; Score 163; DB 1; Length 1854;
Best Local Similarity 97.1%; Pred. No. 1.6e-42;
Matches 166; Conservative 0; Mismatches 5; Indels 0; Gaps 0:

Oy 126 cctagctctcaaaagtctatggccctgtgttcacctgtgttgcctgaaccctag 185
|||
Db 176 CCAATCTCTCAAGGCTATGGCCCTGTGTTCACTCTGTAITTTGGCCTGAACCCATAG 235

Oy 186 tggctgctgatgatatgaagcagtgaaagcctgattgattgattgagagagatttt 245
|||
Db 236 TGGTGTGTCATGGATGAAGCAGTGAAGGAGCCCTGATTGATCTTGGAGAGAGATTT 295

Oy 246 ctggaagaggcattttccctcactggctgaagagctaaacagagatttgta 296
|||
Db 296 CTGGAAGAGGCAATTTCCCACTGGCTGAAGAGCTAACAGAGGATTGGAA 345

RESULT 7
US-08-238-821B-4
: Sequence 4, Application US/08238821B
: Patent No. 5912120
: GENERAL INFORMATION:
: APPLICANT: GOLDSTEIN, Joyce A.
: APPLICANT: ROMKES-SPARKS, Marjorie
: APPLICANT: DE MORAIS, Sonia M.F.
: TITLE OF INVENTION: CLONING, EXPRESSION AND DIAGNOSIS OF HUMAN
: TITLE OF INVENTION: CYTOCHROME P450 2C19: THE PRINCIPAL DETERMINANT OF S-
: TITLE OF INVENTION: MEPHENYTOIN METABOLISM
: NUMBER OF SEQUENCES: 61
: CORRESPONDENCE ADDRESS:

```

```

: ADDRESSEE: Townsend and Townsend and Crew LLP
: STREET: Two Embarcadero Center, 8th Floor
: CITY: San Francisco
: STATE: California
: COUNTRY: US
: ZIP: 94111
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/238,821B
: FILING DATE: 06-MAY-1994
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/201,118
: FILING DATE: 22-FEB-1994
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/864,962
: FILING DATE: 09-APR-1992
: ATTORNEY/AGENT INFORMATION:
: NAME: Liebeschuetz, Joe
: REGISTRATION NUMBER: 37,505
: REFERENCE/DOCKET NUMBER: 15280-19211005
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (650) 326-2400
: TELEFAX: (650) 326-2422
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1854 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: NAME/KEY: Region
: LOCATION: 1..12
: OTHER INFORMATION: /note= "Corresponds to positions -12 to-1
: OTHER INFORMATION: for 25 of Figure 2."
: US-08-238-821B-4

Query Match 21.9%; Score 163; DB 2; Length 1854;
Best Local Similarity 97.1%; Pred. No. 1.6e-42;
Matches 166; Conservative 0; Mismatches 5; Indels 0; Gaps 0:

Oy 126 cctagctctcaaaagtctatggccctgtgttcacctgtgttgcctgaaccctag 185
|||
Db 176 CCAATCTCTCAAGGCTATGGCCCTGTGTTCACTCTGTAITTTGGCCTGAACCCATAG 235

Oy 186 tggctgctgatgatatgaagcagtgaaagcctgattgattgattgagagagatttt 245
|||
Db 236 TGGTGTGTCATGGATGAAGCAGTGAAGGAGCCCTGATTGATCTTGGAGAGAGATTT 295

Oy 246 ctggaagaggcattttccctcactggctgaagagctaaacagagatttgta 296
|||
Db 296 CTGGAAGAGGCAATTTCCCACTGGCTGAAGAGCTAACAGAGGATTGGAA 346

RESULT 8
PCT-US95-05744-4
: Sequence 4, Application PC/TUS9505744
: GENERAL INFORMATION:
: APPLICANT: GOLDSTEIN, Joyce A.
: APPLICANT: ROMKES-SPARKS, Marjorie
: APPLICANT: DE MORAIS, Sonia M.F.
: TITLE OF INVENTION: CLONING, EXPRESSION AND DIAGNOSIS OF HUMAN
: TITLE OF INVENTION: CYTOCHROME P450 2C19: THE PRINCIPAL DETERMINANT
: TITLE OF INVENTION: OF S-MEPHNYTOIN METABOLISM
: NUMBER OF SEQUENCES: 61
: CORRESPONDENCE ADDRESS:

```

ADDRESSEE: Townsend and Townsend Khourie and Crew
STREET: 379 Lytton Avenue
CITY: Palo Alto
STATE: California
COUNTRY: US
ZIP: 94301
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/05744
FILING DATE: 06-MAY-1994
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/238,821
FILING DATE: 06-MAY-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/201,118
FILING DATE: 22-FEB-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/864,962
FILING DATE: 09-APR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Dow, Karen B.
REGISTRATION NUMBER: 29,684
REFERENCE/DOCKET NUMBER: 15280-192-1-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 326-2400
TELEFAX: (415) 326-2422
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1854 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
PCT-US95-05744-4

Query Match 21.9%; Score 163; DB 5; Length 1854;
Best Local Similarity 97.1%; Pred. No. 1.6e-42;
Matches 166; Conservative 0; Mismatches 5; Indels 0; Gaps 0:

Qy 126 cctagctcaaaagctgtatgacctgtgttcactctgtatgttgcctgaaccataag 185
|||||
Db 176 CCAATCTCTCAAAGGCTATGGCCCTGTGTCTACTGTGTATTTGGCTCGAAACCCATAG 235
|||||

Qy 186 tggctgctgcatgatatgaagcagtgtaaggaagccctgattgattcttgaggagagtttt 245
|||||
Db 236 TGGTGTGCTGTCATGATATGAAGCAGTGAAGGAAGCCCTGATTGATCTTGGAGAGAGTTT 295
|||||

Qy 246 ctggaagaggcattttccctcctggctgaagagcctaacaagagatttgta 296
|||||
Db 296 CTGGAAGAGGCATTTTCCCACTGGCTGAAAGAGCTAAAGAGAGATTGGAA 346
|||||

RESULT 9
US-08-201-118-14
Sequence 14, Application US/08201118
Patent No. 5786191
GENERAL INFORMATION:
APPLICANT: GOLDSTEIN, Joyce A.
APPLICANT: ROMKES-SPARKS, Marjorie
TITLE OF INVENTION: CLONING AND EXPRESSION OF COMPLEMENTARY
TITLE OF INVENTION: DNAs FOR MULTIPLE MEMBERS OF THE HUMAN CYTOCHROME P450 2C
TITLE OF INVENTION: SUBFAMILY
NUMBER OF SEQUENCES: 44
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend Khourie and Crew
STREET: 379 Lytton Avenue
CITY: Palo Alto

STATE: California
COUNTRY: US
ZIP: 94301
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/201.118
FILING DATE: 22-FEB-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/864,962
FILING DATE: 09-APR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Liebeschuetz, Joe
REGISTRATION NUMBER: 37,505
REFERENCE/DOCKET NUMBER: 15280-192-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 326-2400
TELEFAX: (415) 326-2422
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 1892 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: YES
US-08-201-118-14

Query Match 20.6%; Score 153.4; DB 1; Length 1892;
Best Local Similarity 91.8%; Pred. No. 2e-39;
Matches 157; Conservative 0; Mismatches 14; Indels 0; Gaps 0:

Qy 126 cctagctctcaaaagctgtatgacctgtgttcactctgtatgttgcctgaaccataag 185
|||||
Db 206 CCAATNTCTCAAAGTCTATGGCCCTGTGTCTACTGTGTATTTGGCTCGAAACCCATAG 265
|||||

Qy 186 tggctgctgcatgatatgaagcagtgtaaggaagccctgattgattcttgaggagagtttt 245
|||||
Db 266 TGGTGTGCTGTCATGATATGAAGCAGTGAAGGAAGCCCTGATTGATCTTGGAGAGAGTTT 325
|||||

Qy 246 ctggaagaggcattttccctcctggctgaagagcctaacaagagatttgta 296
|||||
Db 326 CTGGAAGAGGCANTTTCCCACTGGCTGAAAGAGANTAACANAGAGANTTGAA 376
|||||

RESULT 10
US-08-238-821B-14
Sequence 14, Application US/08238821B
Patent No. 5912120
GENERAL INFORMATION:
APPLICANT: GOLDSTEIN, Joyce A.
APPLICANT: ROMKES-SPARKS, Marjorie
APPLICANT: DE MORAIS, Sonia M.F.
TITLE OF INVENTION: CLONING, EXPRESSION AND DIAGNOSIS OF HUMAN
TITLE OF INVENTION: CYTOCHROME P450 2C19: THE PRINCIPAL DETERMINANT OF S-
TITLE OF INVENTION: MEPHENYTOIN METABOLISM
NUMBER OF SEQUENCES: 61
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: California
COUNTRY: US
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

```

: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCI/US95/05744
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/238,821
: FILING DATE: 06-MAY-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/201,118
: FILING DATE: 22-FEB-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/864,962
: FILING DATE: 09-APR-1992
: ATTORNEY/AGENT INFORMATION:
: NAME: DOV, Karen B.
: REGISTRATION NUMBER: 29,584
: REFERENCE/DOCKET NUMBER: 15280-192-1-1
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 326-2400
: TELEFAX: (415) 326-2422
: INFORMATION FOR SEQ ID NO: 14:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1892 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: HYPOTHEetical: YES
: PCI-US95-05744-14

```

	Query Match	20.6%	Score 153.4;	DB 5;	Length 1892;
	Best Local Similarity	91.8%	Pred. No. 2e-39;		
	Matches 157;	Conservative	0;	Mismatches 14;	Indels 0;
	Gaps	0;			
Qy	126	cctagctctcaaaaggtctatggccctctgttccactctgtatttggccctgaaccccatag	185		
Db	206	CCAAINTCTCAAAAGTCTATGGCCCTGTGTTTCACNTGTATT	265		
Qy	186	tggctgcctcgtgatatgaagcagctgaaggaagccctgattgactctggaggagagtttt	245		
Db	266	TGGTGTGTCATGGATATGAAGCAGTGAAGGAAGCCCTGATTGATCNTGGAGGAGGATTTT	325		
Qy	246	ctggaagaggcgattttcccccctggctgaaagagctaaacagagatttggtta	296		
Db	326	CTGCAAGCGGCANTTTCCTCTGGCTGAACAGAGNTAACANAGGANTTTCGAA	376		

RESULT 12
 US-08-201-118-2
 : Sequence 2, Application US/08201118
 : Patent No. 5786191
 : GENERAL INFORMATION:
 : APPLICANT: GOLDSTEIN, Joyce A.
 : APPLICANT: ROMKES-SPARKS, Marjorie
 : TITLE OF INVENTION: CLONING AND EXPRESSION OF COMPLEMENTARY
 : TITLE OF INVENTION: CLONING AND EXPRESSION OF COMPLEMENTARY
 : TITLE OF INVENTION: DNAS FOR MULTIPLE MEMBERS OF THE HUMAN CYTOCHROME P450 C
 : TITLE OF INVENTION: SUBFAMILY
 : NUMBER OF SEQUENCES: 44
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Townsend and Townsend Khourie and Crew
 : STREET: 379 Lytton Avenue
 : CITY: Palo Alto
 : STATE: California
 : COUNTRY: US
 : ZIP: 94301
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: PatentIn Release #1.0, Version #1.25

```

: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/864,962
: FILING DATE: 09-APR-1992
: ATTORNEY/AGENT INFORMATION:
: NAME: Liebeschuetz, Joe
: REGISTRATION NUMBER: 37,505
: REFERENCE/DOCKET NUMBER: 15280-192110US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (650) 326-2400
: TELEFAX: (650) 326-2422
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1746 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: FEATURE:
: NAME/KEY: Region
: LOCATION: 1..5
: OTHER INFORMATION: /note= "Corresponds to positions -5 to-1
: OTHER INFORMATION: for IIA of Figure 2."
: US-08-238-821B-2

Query Match          20.1%   Score 149.6; DB 2; Length 1746;
Best Local Similarity 97.4%; Pred. No. 3.2e-38;
Matches 152; Conservative 0; Mismatches 4; Indels 0; Gaps

Qy 461 ttgggaattgttttcagcaatgaaagaataaggagagatccgcgtttctcctcatg 521
    |||||
Db 333 TTGGAATCGTTTTCAGCAATTGGAACAGATGGAAGCAGATCCGGCGCTTTCTCCCTCATG 393

Qy 521 acgctcggaatttgggatgggagagagagacatggagaccgtgttcaagaggaagcc 581
    |||||
Db 393 ACCTCGCGAAATTGGGATGGGAGAGAGAGCATTGAGACCCTGTTCGAAGAGAAGCC 451

Qy 581 cgctgccttgaggagagttagaaaaaaccaagggt 616
    |||||
Db 453 CGCTGCCTTGCGAGGAGTTGAGAAAAAACCAAGGCT 488

RESULT 14
PCT-US95-05744-2
: Sequence 2, Application PC/TUS9505744
: GENERAL INFORMATION:
: APPLICANT: GOLDSTEIN Joyce A.
: APPLICANT: ROMKES-SPARKS, Marjorie
: APPLICANT: DE MORAIS, Sonia M.F.
: TITLE OF INVENTION: CLONING, EXPRESSION AND DIAGNOSIS OF HUMAN
: TITLE OF INVENTION: CYTOCHROME P450 2C19: THE PRINCIPAL DETERMINANT
: TITLE OF INVENTION: OF S-MEPHYNYTOIN METABOLISM
: NUMBER OF SEQUENCES: 61
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Townsend and Townsend Kourie and Crew
: STREET: 379 Lytton Avenue
: CITY: Palo Alto
: STATE: California
: COUNTRY: US
: ZIP: 94301
: COMPUTER READABLE FORM: disk
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US95/05744
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/238,821
: FILING DATE: 06-MAY-1994
: PRIOR APPLICATION DATA:

```

